



# Génomique évolutive de l'agent pathogène de la tavelure du pommier, *Venturia inaequalis*, dans le cadre de la domestication de son hôte

Marie de Gracia

## ► To cite this version:

Marie de Gracia. Génomique évolutive de l'agent pathogène de la tavelure du pommier, *Venturia inaequalis*, dans le cadre de la domestication de son hôte. Sciences agricoles. Université d'Angers, 2014. Français. NNT : 2014ANGE0010 . tel-01304652

**HAL Id: tel-01304652**

**<https://theses.hal.science/tel-01304652>**

Submitted on 20 Apr 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Thèse de Doctorat

**Marie DE GRACIA**

*Mémoire présenté en vue de l'obtention du  
**grade de Docteur de l'Université d'Angers**  
sous le label de L'Université Nantes Angers Le Mans*

**École doctorale :** Végétal, Environnement, Nutrition,  
Agroalimentaire, Mer (VENAM)

**Discipline :** 3DOSa

**Spécialité :** Sciences agronomiques

**Unité de recherche :** EcoFun, IRHS (Université d'Angers  
Inra, Agrocampus-Ouest) - UMR 1345

**Soutenue le** 18 décembre 2014

**Thèse N° :** 1440

**Génomique évolutive de l'agent pathogène de  
la tavelure du pommier, *Venturia inaequalis*,  
dans le cadre de la domestication de son hôte.**

## JURY

**Rapporteurs :** Elisabeth FOURNIER, Chargée de recherche, INRA  
François DELMOTTE, Directeur de recherche, INRA

**Examineurs :** Lilian GOUT, Maître de conférences, AgroParisTech  
Nicolas BIERNE, Directeur de recherche, CNRS

**Directeur de Thèse :** Bruno LE CAM, Directeur de recherche, INRA

**Encadrant de Thèse :** Christophe LEMAIRE, Maître de conférences, Université d'Angers





## Remerciements

L'heure des remerciements, enfin! Exercice moins facile que je ne le pensais. Difficile de résumer ces 6 dernières années sans oublier quelqu'un. Et oui car moi, quand je suis rentrée à l'INRA, j'avais encore des couettes et mon paquet de BN dans la poche :) . Je vais essayer de faire des remerciements dignes d'un Ecofunien, du genre à la N. Anger (référence ultime du remerciement clair et précis).

Je me souviens du premier jour où j'ai posé le pied ici ... l'odeur de mycose dans les couloirs est toujours présente ... tout comme l'ambiance qui règne dans notre petite équipe qui commence à devenir grande.

En premier lieu, je tiens à remercier les patrons, les padres :  
Mon directeur de thèse : Bruno Le Cam, Brubo, pour m'avoir offert toutes ces opportunités de travailler au sein d'Ecofun depuis ma L3. J'ai pu aborder plusieurs thématiques au sein du labo. J'ai bien conscience de la chance que j'ai eue d'avoir un directeur de thèse s'investissant autant pour son équipe. Je te remercie pour ta disponibilité et ta confiance.

Mon encadrant de thèse : Christophe Lemaire, Krostif, qui m'a repêchée de mon parcours «Triton, dessin & co». Tu m'as appris une grande partie de ce que je peux savoir. Des raboutages aux « N » près à nos premiers scripts de la mort qui tue, j'ai un sacré nombre de souvenirs. J'espère ne pas t'avoir déçu et MERCI pour tout. C'est bon je pars, tu vas pouvoir enfin récupérer un bureau digne de ce nom. En prime je te laisse le rat mort qui loge dans mon bureau depuis le début de la rédaction.

Je vous remercie tous les deux d'avoir été là jusqu'au bout et durant tous mes moments de doute. De m'avoir supportée. Je sais que si vous n'aviez pas été là, je ne serai pas en train d'écrire cette page à l'heure actuelle.

Aux membres de mon comité qui ont pu apporter de nombreux conseils et ajustements à mon travail : Julien Dutheil, Stéphane de Mita et Joëlle Amselem. Merci Julien pour l'aide apportée pour la réalisation de mon Hapmap dont j'ai pu rêver des nuits et des nuits, même si cela n'a pas abouti. Tu as toujours été disponible pour moi et répondu à toutes mes questions. Merci Stéphane pour tes informations pour Egglib et tes encouragements jusqu'à la fin qui m'ont remonté le moral je dois l'avouer.

Je tiens également à remercier Elisabeth Fournier et François Delmotte d'avoir accepté d'être rapporteurs de ma thèse, ainsi que Lilian Gout et Nicolas Bierne d'être mes examinateurs.

Durant ces 6 ans à l'INRA, j'ai pu côtoyer de nombreuses personnes durant les colloques et congrès, et en particulier au REID et à Aussois. A cet instant, je pense en particulier à 2 personnes qui ont tout fait changer. Et qui donnent envie de rester dans ce milieu.

Merci à Dieu, comme je l'appelle ici, Jérôme Gouzy. Cette semaine à Toulouse, te voir coder aussi vite que ton ombre, ta disponibilité depuis 2 ans, tout le boulot que tu as pu fournir, Wouahou. Tu m'as initiée à mes premières lignes de code et ce n'était vraiment pas gagné !

Camille Roux, et nos discussions Skype ! Une aprèm avec toi équivalait à 1 semaine de boulot toute seule. Avec toi, tout est toujours plus simple et accessible. Mon seul regret : de ne pas t'avoir connu plus tôt.

Je remercie bien entendu l'ensemble des membres et anciens membres de l'équipe Ecofun que j'ai côtoyés : Marie-Noëlle Bellanger, Pascale Expert, Frédérique Didelot, Natacha Sapoukhina, Valérie Caffier, Julie Mintsas, Baba Issa Camara, Amandine Le Van.

Merci aux nombreuses personnes du bâtiment C, et en particulier Sophie Cesbron, avec qui j'ai partagé un petit peu de temps son bureau, Sylvain Gaillard et Martial Briand pour votre disponibilité à chaque fois que je viens vous déranger pour une question bioinfo.

Merci à tous les jeunes padawans avec qui j'ai passé beaucoup de temps ces dernières années. Brice (cette première fois où tu m'as vue sous la hotte : Qu'est-ce que tu me veux toi !), Matthieu Gaucher, Jeff Guimbaud, Arnaud Indiana (désolée pour toutes les questions :D ), Matthieu Jourdan et Céline Rousseau (le sosie mdr), Amandine Cuntz (et son rire ahah) & Déborah Merda (je t'invite à mes 27 ans, sois à l'heure cette fois-ci) pour leur soutien ces derniers mois. Merci les filles ! Thibault Leroy, merci d'avoir toujours été là pour moi depuis le début. Diane Leforestier, wesh mec. Je ne sais pas quoi choisir entre ma réputation que tu as accentuée ou tes gâteaux ! Merci et je t'apporte les papiers pour l'entreprise dans 6 mois !

Au maître des clés, grand maître Chen : de notre premier match de foot à nos tisanes du dimanche soir, les moments passés en ta compagnie me tiennent beaucoup à cœur.

Merci ma Coco, ma sauveuse ! Malgré le fait que je t'ai dégoûtée à vie des extractions, j'espère t'avoir laissé des bons moments (cf : magnifique photo dans le labo). J'ai apprécié tous les mois qui ont défilé en ta compagnie. Tu as été une de mes plus belles rencontres de ces dernières années. Bon je te fais toujours la tête pour m'avoir laissée toute seule pendant 6 mois.

Merci Calmes ;) La cohabitation n'a pas toujours été facile et évidente au début. Après t'avoir calmé un peu (nos réputations de super connards oblige!), j'ai découvert quelqu'un d'attachant, avec un caractère de m\*\*\*\* comme moi, et avec le même rire ... qui fait mal aux oreilles. Tu as été là et m'as motivée jusqu'au

bout. Merci pour tout copain. « J'ai mis ma blouse, il fait trop froid »  
#BrèvesDePaillasse

Gaëlle Laloi alias Gaya, Li et future Shirley (et oui j'ai décidé de ton surnom pour ta troisième année, un mélange de junky qui n'est pas sorti des années 90 et collectionneuse d'images panini). J'aurai aimé partager plus d'années de thèse avec toi, et mon pop-corn. Bon courage pour la suite, mais on aura encore de nombreuses occasions de se voir.

Merci mes loulous, Mathilde Cascales et Wesley Delage. Roukine, j'ai appréhendé l'encadrement de ton stage. Pas facile de donner des ordres quand on a le même âge. Tu t'en es super bien sortie. Bon on ne fera jamais cette super coloc de la mort qui tue « De Laf, Cascales, De Gracia : viva la paella ! ».

Merci Biket. Avec un grand B :). Je pense que tout est dit ! <3

Je pense bien évidemment à la famille, sisi la famille, Alek (tellement contente que tu sois là pour la fin), mon Franfran, les Dupuis (cette thèse marque la fin du binôme, et 11 ans de collaboration, de toute façon, cela ne s'appelle plus gérer à ce point là, cela s'appelle Dupuisdegracier), Damien, Alice, Maidie, mon Kinder (d'être mon Kinder et parce qu'il y a des choses que ne changeront jamais. K&M<3).

Une petite pensée pour la BRAF, Posso représente !

Merci à la Famille Fonfon pour tout ce que vous avez fait, depuis toujours.

Mes sponsors : H&V, Nutella, les lamas, le site du zéro, Lipton, la ligue des champions, l'huile d'olive, Leva's Polka, de Laf, Panzani, Les Ducs, « INTERNET mon amour, mon royaume ».

A toi ma Jacqueline. Désolée de ne pas t'avoir élevée plus haut, mais comme tu le sais, on ne joue pas tous avec les mêmes jojo's.

Un grand merci à ma famille; mes parents, qui m'ont maintenue en vie ces dernières années. A minimoi.

Merci à la double paire d'yeux bleus que j'ai la chance d'apercevoir à chaque réveil <3 .

# Stroumpfette et # Caravane

PS : je n'ai jamais apprécié l'humour d'Ecofun en fait :D



*A Dada et Mimi*





## Table des matières

<b>Introduction bibliographique</b>	13
<b>1. Le processus de domestication : L'impact des agrosystèmes sur les pathogènes associés aux hôtes domestiqués</b>	13
1.1. La domestication	13
1.1.1. Contexte général – Définition	13
1.1.2. Conséquences évolutives de la domestication chez les hôtes	18
1.2. L'évolution de l'agriculture et la diffusion intensive des agents pathogènes	21
1.2.1. Co-évolution des agents pathogènes et des hôtes dans les agrosystèmes	21
1.2.2. Les mécanismes évolutifs d'émergence des pathogènes : hypothèses écologiques	23
1.2.3. Hypothèses non écologiques : les incompatibilités Dobzhansky-Muller	24
1.2.4. Les contacts secondaires chez les agents phytopathogènes	26
1.3. Diversification post-domestication : conséquences évolutives vis à vis de l'agent phytopathogène	27
<b>2. Le pathosystème <i>Venturia inaequalis</i> / pommier</b>	35
2.1. La tavelure du pommier	35
2.1.1. Importance économique	35
2.1.2. Symptomatologie et cycle de vie	36
2.2. Processus de domestication et histoire évolutive	38
2.2.1. Origine et domestication du pommier cultivé	38
2.2.2. Histoire évolutive des populations de <i>Venturia inaequalis</i> au sein du centre de domestication du pommier	38
a) <i>Structure des populations de Venturia inaequalis au Kazakhstan</i>	40
b) <i>Estimation du temps de divergence</i>	40
c) <i>Evolution des traits d'histoire de vie de la tavelure</i>	41
d) <i>La résistance Vf</i>	42
2.3. Questions de recherche	43



<b>Résultats</b>	45
<b><u>Chapitre 1</u></b>	47
« Origine de la virulence Vf chez la tavelure du pommier »	
Introduction	51
Materials & Methods	55
Results	58
Discussion	64
References	69
<b><u>Chapitre 2</u></b>	75
« Evolution des traits d’histoire de vie de <i>Venturia inaequalis</i> en lien avec la domestication du pommier »	
Introduction	78
Materials and Methods	83
Results	85
Discussion	90
References	95
<b><u>Chapitre 3</u></b>	101
« Signatures génomiques de l’impact de la domestication du pommier sur son pathogène <i>Venturia inaequalis</i> »	
Introduction	103
Materials & Methods	108
Results	113
Discussion	124
References	129
<b>Discussion générale</b>	135
<b>Bibliographie</b>	143

# **Introduction bibliographique**

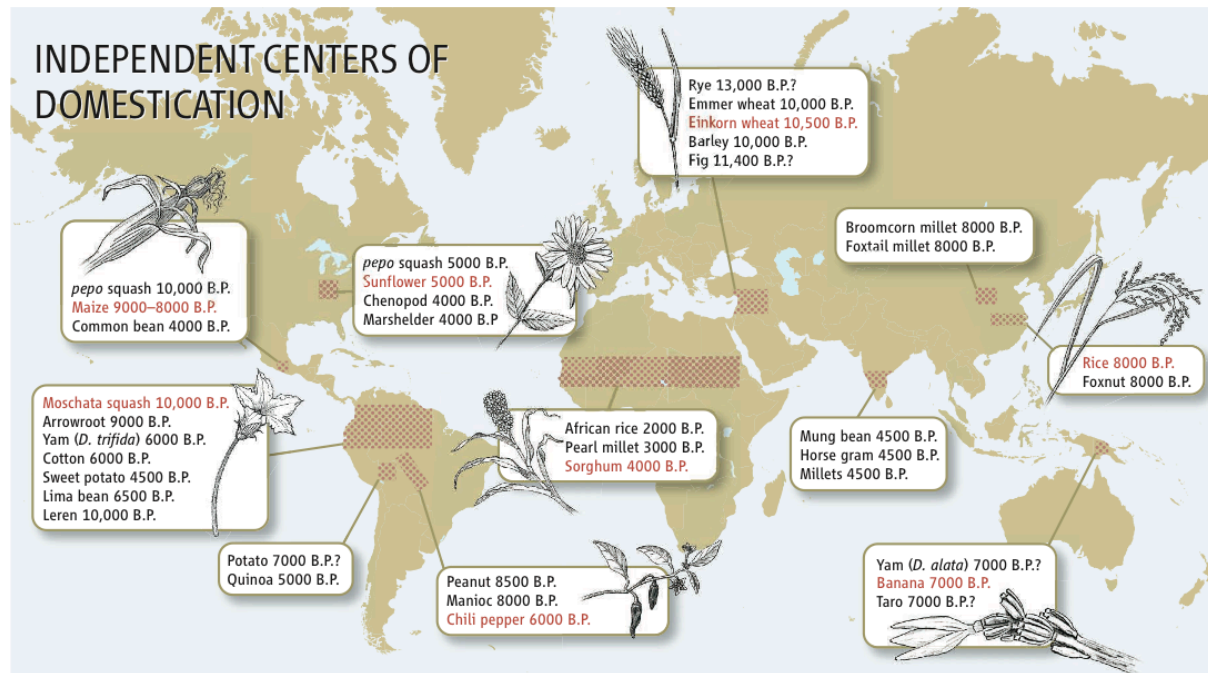
---

# **1. Le processus de domestication : L'impact des agrosystèmes sur les pathogènes associés aux hôtes domestiqués**

## **1.1. La domestication**

### **1.1.1. Contexte général – Définition**

Les modalités par lesquelles les espèces s'adaptent à leur environnement sont au cœur de la théorie de l'évolution décrite dans l'Origine des espèces (Darwin, 1859). La métaphore de la domestication a permis à Darwin d'étayer son concept de variation par sélection (Darwin, 1882). En effet, la domestication peut être considérée comme étant un processus rapide d'évolution par sélection récurrente dirigée par l'action humaine. La domestication des espèces animales et végétales peut être considérée comme une étape déterminante dans l'apparition de l'agriculture. Par sélection des phénotypes d'intérêt agronomique et croisements favorisés entre ces phénotypes, les premiers agriculteurs sont parvenus à s'assurer une disponibilité régulière en ressources alimentaires animales et végétales. Ainsi les débuts de l'agriculture ont marqué une profonde transformation des espèces, par la création de combinaisons phénotypiques (et donc génotypiques) inédites. La domestication des espèces végétales et animales est l'un des changements les plus marquants des derniers millénaires de l'histoire humaine. Par son association directe avec le développement de l'agriculture, il s'agit de l'un des tournants décisifs de notre histoire tant du point de vue économique, social et culturel ; la domestication a régi notre alimentation actuelle, l'essor des civilisations ainsi que notre démographie mondiale. L'Homme a commencé à domestiquer certaines espèces animales et végétales au Néolithique il y a environ 13 000 ans : période décisive de l'évolution de la condition humaine caractérisée par la transition de sociétés de chasseurs-cueilleurs à des sociétés agricoles sédentaires (Diamond, 2002). Les premiers centres de domestication se localisent au niveau du croissant fertile mais la domestication des plantes s'est faite de manière indépendante les unes des autres dans le monde (Figure 1).



**Figure 1. Centres de domestication des plantes (Balter, 2007)**

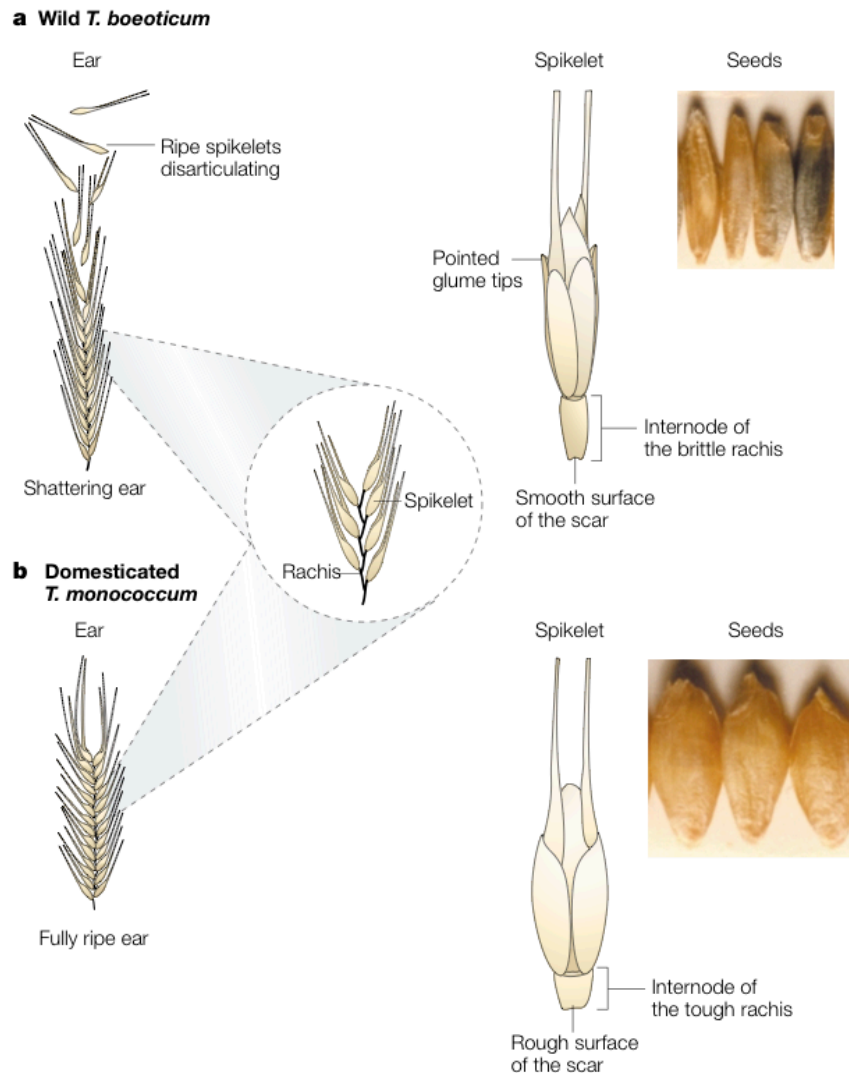
La domestication peut se définir comme le stade final d'un processus continu et dynamique, qui commence par l'exploitation en milieu naturel de phénotypes sauvages et la sélection progressive d'individus à intérêt agronomique jusqu'à leur culture ou élevage en milieu artificiel (Pickersgill, 2007).

D'après cette définition, l'action de domestiquer a pour but originel de satisfaire les besoins humains. Ces besoins peuvent se regrouper en deux catégories : la protection face aux prédateurs, ou ravageurs et l'alimentation. Dans la première catégorie, la domestication du chien représente un exemple classique. Les chiens ont très tôt permis la protection des humains, des troupeaux contre les prédateurs, et de tout autre bien contre toute sorte d'agression. Toutefois, une autre catégorie a émergé comme l'atteste la domestication du chat. Contrairement aux autres espèces domestiquées, la domestication de ce dernier résulterait d'un processus culturel dans un premier temps, puis économique. En effet, la vénération des chats par les égyptiens impliquait de les garder à proximité des temples et maisons (Malek, 1993). Néanmoins, ses qualités de chasseur pour lutter contre l'invasion des rongeurs présents dans les réserves de blé des villages de la vallée du Nil étaient fortement appréciées. Au fil du temps, les rencontres occasionnelles entre les hommes et les chats se seraient transformées en contacts plus intimes, poussant les humains à le domestiquer afin de limiter les populations d'animaux nuisibles pour les habitations et les réserves de nourriture (Serpell, 2000).

Cependant la domestication n'est pas un processus irréversible. En effet, certaines espèces peuvent, dans certains cas, retourner à l'état sauvage. Le chien ou le vison d'Amérique sont des exemples qui caractérisent ce phénomène de retour à l'état sauvage appelé feralisation. Toutefois, lors de domestications extrêmes comme celle du Bombyx du mûrier (*Bombyx mori*) pour la production de soie, le retour à l'état sauvage s'avère difficile voire impossible. Ce papillon a été domestiqué en Chine il y a 5 000 ans. Sa survie et sa reproduction sont entièrement contrôlées par la volonté de l'Homme. Contrairement à l'exemple précédent, le papillon risquerait de disparaître si la production de soie venait à s'arrêter.

En dehors de ces cas extrêmes, le processus de domestication peut dans certains cas se caractériser par l'existence d'un *syndrome de domestication* c'est à dire par la fixation de différences phénotypiques et génétiques distinguant les taxons sauvages et domestiqués (Pickersgill, 2007). Par extension, on nomme syndrome de domestication les transitions morphologiques, phénologiques mais également physiologiques entre les formes sauvages et cultivées. Ces transitions sont associées à la forte sélection anthropique des traits utiles : liés à la production des graines ou aux conditions de récolte. A titre d'exemple, la comparaison morphologique entre l'en grain qui est le premier blé domestiqué (*Triticum monococcum*) et sa forme sauvage (*Triticum boeoticum*) permet de visualiser les impacts de la domestication (Figure 2). L'en grain sauvage a de petites graines, tandis que celles de la forme domestiquée sont plus grandes. Les épillets domestiques ont un rachis plus solide contenant les graines et possèdent des glumes qui ne perturbent pas la bonne dispersion des graines contrairement aux formes sauvages (Salamini *et al.*, 2002).

Outre le fait que la domestication induit des changements phénotypiques et génétiques considérables chez les plantes et les animaux, son étude permet également d'obtenir une masse d'informations quant à la démographie de l'époque, la transformation de l'agriculture et de manière plus générale, de l'Histoire des civilisations. C'est ainsi qu'on observe souvent une combinaison entre les données archéologiques et génétiques pour déterminer plus précisément les périodes de domestication.



**Figure 2. Différences morphologiques entre forme sauvage et domestiquée : cas de la domestication de l'engrain (Salamini *et al.*, 2002).**

L'engrain *Triticum monococcum* (b) et son ancêtre *Triticum boeoticum* (a) se distinguent principalement par la taille des graines et la forme de l'épillets.

La domestication a donc été une phase déterminante dans l'essor des sociétés humaines en permettant non seulement le développement de l'agriculture mais aussi l'apparition de nouveaux aliments tels que la bière et le pain au levain, issus de l'utilisation de souches de levures *Saccharomyces cerevisiae*. L'action de la sélection récurrente entreprise par nos ancêtres a eu plusieurs conséquences sur les traits d'histoire de vie des organismes domestiqués. Par la suite nous nous intéresserons aux conséquences évolutives de la domestication chez les plantes et notamment sur les agents pathogènes infectant ces organismes.

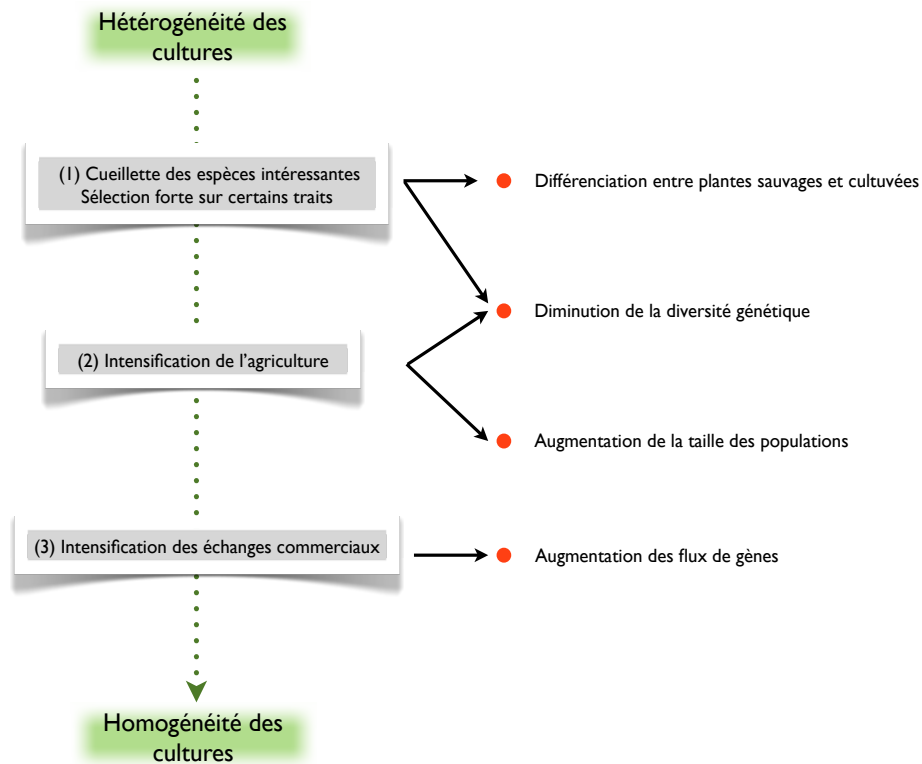
### 1.1.2. Conséquences évolutives de la domestication chez les hôtes

Avant d'énoncer les conséquences évolutives possibles de la domestication sur les agents pathogènes, il faut dans un premier temps aborder les principales modifications engendrées par la domestication chez les hôtes.

La domestication des plantes et le développement de l'agriculture ont conjointement modifié de manière significative la structure de l'environnement. La transition vers l'agriculture s'est traduite par la densification des espèces exploitées sur des espaces limités : les parcelles. La distribution des plantes cultivées dans l'espace s'est ainsi trouvée modifiée (Figure 4). L'intensification de l'agriculture a provoqué une augmentation de la taille des populations et une homogénéité génétique des espèces cultivées (Figure 3). En effet, la forte sélection exercée sur les organismes selon des caractères jugés favorables par l'homme s'est accompagnée de forts goulets d'étranglement réduisant ainsi significativement la diversité génétique des espèces domestiquées (Londo *et al.*, 2006). L'exemple le plus révélateur est le blé domestique qui ne présente que 30 % de la variabilité ancestrale. Communément, cette diminution se situe plutôt autour des deux tiers de la diversité nucléotidique ancestrale, comme chez le maïs ou l'avoine (Buckler *et al.*, 2001).

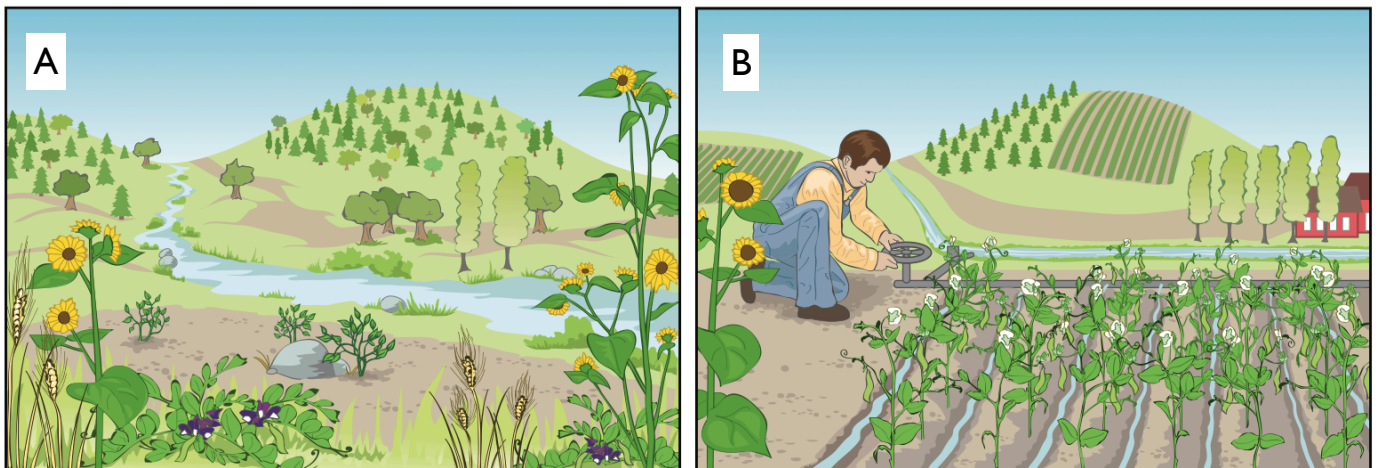
De plus, les changements de pratiques culturales affectent la configuration des organismes les uns par rapport aux autres. Les agrosystèmes constituent alors des structures particulières et sont presque comparables à un milieu « clonal ». L'environnement de vie qu'ils proposent est alors très peu affecté par les changements de conditions environnementales. En effet, les plantes cultivées poussent dans des habitats favorisés par l'action anthropique. Ces milieux sont le plus souvent fertilisés, irrigués et protégés de toute prédation ou de compétition interspécifique. A ce titre, comme cela a été mentionné plus haut, la valeur adaptative des espèces domestiquées et cultivées n'est pas optimale en milieu sauvage.

Enfin, l'intensification de l'agriculture est en lien direct avec l'intensification des échanges commerciaux via la mondialisation (Figure 3). Le principal résultat est l'augmentation des flux de gènes grâce à la propagation mondiale par les activités humaines. Ces échanges permettent de maintenir la diversité génétique au niveau mondial, alors que celle-ci tend à diminuer localement. La tendance à la diminution de la diversité génétique par rapport aux espèces mères sauvages a particulièrement été observée chez les taxons de graminées tels que le blé (Kilian *et al.*, 2007), le riz (Caicedo *et al.*, 2007) et le maïs (Wright *et al.*, 2005) (voir tableau 2 in Glémin et Bataillon, 2009)



**Figure 3. Evolution des hôtes : transition d'un milieu sauvage à un agrosystème**

Les encadrés en gris représentent les 3 principales étapes du passage d'un milieu sauvage à un milieu domestiqué. Les points rouges indiquent les conséquences qui en découlent.



**Figure 4. Principales différences visibles entre les milieux sauvages (A) et les agrosystèmes (B) (Stukenbrock & Macdonald, 2008)**

La comparaison d'un écosystème naturel (A) et sauvage (B) se caractérise par une diminution de la diversité des espèces, une homogénéité de l'environnement ainsi qu'une densité accrue dans le compartiment domestique.



Chez différentes espèces de graminées, la diversité nucléotidique des plantes domestiquées représente environ le tiers de la diversité observée sur plantes sauvages (Glémin & Bataillon, 2009). Cette réduction de diversité peut s'expliquer facilement par les différents épisodes de sélection artificielle qui ont eu pour effet de réduire considérablement les tailles efficaces et par là même, d'accroître l'intensité de la dérive au sein des populations cultivées. Une telle baisse de diversité a sans doute permis la fixation de caractères délétères dans la nature, permettant l'établissement du syndrome de domestication : la fixation de traits morphologiques, ou d'histoire de vie, différenciant les espèces domestiquées de leurs ancêtres sauvages. Les modifications des modalités de dispersion des graines allant de la dispersion libre chez les espèces sauvages à une dispersion plus dépendante de l'action humaine, sont les principales marques du syndrome de domestication. Les graines ou les fruits sont, en effet, les premières ressources alimentaires d'origine végétale, de par l'apport énergétique de l'amidon et autres sucres. Pour autant, si l'évolution des caractères d'intérêt agronomique comme la taille des graines ou des fruits est bien étudiée, peu de données rendent compte de l'évolution de la résistance aux pathogènes au cours de la domestication. Il est peu probable que la résistance aux pathogènes fut d'emblée sélectionnée par les premiers agriculteurs. On peut donc dériver deux types d'hypothèse : soit les résistances seraient apparues suite aux différentes sélections artificielles, par adaptation *de novo* ; soit les résistances étaient déjà présentes dans les populations sauvages et ont été incidemment sélectionnées lors de la domestication.

Enfin, le passage du milieu sauvage au milieu agricole et l'évolution des pratiques culturales tels que l'irrigation, l'amendement, la taille, pour ne citer qu'eux, ont pu avoir un impact sur les relations entre la plante et ses parasites. En effet, en considérant que les traits d'histoire de vie du parasite sont adaptés à la phénologie de l'hôte, nous pouvons nous interroger sur les conséquences du syndrome de domestication chez la plante sur les traits d'histoire de vie de ses pathogènes. Dans la suite de cette introduction nous tenterons de développer les conséquences attendues du passage d'un environnement sauvage à un agrosystème pour un organisme pathogène.

## **1.2. L'évolution de l'agriculture et la diffusion intensive des agents pathogènes**

En raison de leur interdépendance, les pressions de sélection qui interviennent chez l'hôte cultivé influencent directement le cortège de pathogènes associés à ce dernier. Pour les agents pathogènes se dispersant à faible distance, l'aire de distribution correspond généralement à la densité d'hôte dans un environnement. Toutefois cette aire peut être perturbée par les activités humaines comme la rotation des cultures, l'extension vers de nouvelles aires agricoles, ou *a contrario* la déprise agricole. Loin de reproduire les environnements sauvages ancestraux, les agroécosystèmes offrent plutôt des habitats à fortes ressources trophiques et à faible compétition interspécifique. Cependant, ces habitats sont instables sur des échelles temporelles plus larges. L'évolution des pratiques agricoles des débuts de l'agriculture il y a 12 000 ans à l'industrialisation agricole depuis la fin de la seconde guerre mondiale dans les pays développés ne constitue-t-elle pas un véritable bouleversement pour les habitats des parasites de plantes cultivées? Comment les populations de pathogène co-évoluant avec leurs hôtes se sont adaptées à ces nouveaux environnements anthropisés ? Comment, en outre, les activités humaines ont-elles renforcé la diffusion des agents pathogènes via la domestication mondiale des plantes, l'importation de nouvelles cultures et l'intensification des échanges commerciaux ? Nous adressons ici la question de la dynamique adaptative des pathogènes au travers du passage d'un habitat sauvage à un habitat anthropisé. La compréhension des processus évolutifs menant à l'émergence de nouveaux traits d'histoire de vie chez les pathogènes est essentielle afin d'évaluer les risques posés par des agents pathogènes émergents et ainsi d'améliorer le développement et la mise en œuvre de stratégies de lutte durable.

### **1.2.1. Co-évolution des agents pathogènes et des hôtes dans les agrosystèmes**

Dès l'apparition de l'agriculture pendant le Néolithique, les humains ont dû faire face aux nombreux dégâts causés par les parasites (Balter, 2007 ; Stukenbrock & MacDonald, 2008) : insectes herbivores, agents pathogènes, microorganismes. Cette co-évolution plantes-

pathogènes est un problème capital à étudier, notamment au travers du processus de domestication où l'émergence peut être favorisée par l'intensification de l'agriculture, l'homogénéité des cultures et l'augmentation des échanges (Anderson *et al.*, 2004). Les modifications environnementales sont généralement considérées comme la cause première de la diversification des pathogènes. La théorie de la reine rouge décrit les relations biotiques entre le couple hôte-parasite (Van Valen, 1973). Cette théorie suggère une course évolutive constante entre les espèces qui sont en interaction. Hôte et parasite s'imposent des pressions de sélection antagonistes l'un par rapport à l'autre. Ainsi, toute modification chez l'hôte et, par extension, de l'environnement va avoir un impact direct sur les populations de pathogènes associés et réciproquement. Cette co-évolution est alimentée par l'apparition de résistances chez l'hôte et de mécanismes de contournement chez les agents pathogènes. Dans le contexte de la domestication et/ou du passage aux agroécosystèmes, l'influence des pressions de sélection exercées par l'homme sur certains traits d'histoire de vie de l'hôte et la modification de l'environnement biotique et abiotique doit théoriquement avoir un impact fort sur l'évolution des pathogènes. Ainsi, peut-on supposer que les changements évolutifs sont plus rapides dans les agrosystèmes que dans les écosystèmes naturels (cf § 1.3) (Stukenbrock & MacDonald, 2008) ?

De ce point de vue, l'étude des conséquences de la domestication sur les agents pathogènes présents lors de la domestication des hôtes, devient une problématique tout aussi importante que l'analyse de l'impact de la sélection artificielle sur les plantes elles mêmes. Après avoir fait un point sur les différents mécanismes d'émergence, nous allons nous intéresser aux conséquences majeures de la domestication sur les agents phytopathogènes et souligner l'intérêt de prendre en compte l'histoire évolutive des populations de pathogènes pour mieux comprendre leurs dynamiques adaptatives.

L'évolution des pathogènes, et notamment des organismes phytopathogènes, est le plus souvent décrite au travers d'un prisme adaptatif lié au modèle gène pour gène (Flor, 1971). Pour autant, nous tenterons de montrer ici que d'autres mécanismes non purement adaptatifs, donc liés à l'environnement peuvent intervenir sur la dynamique d'émergence et d'évolution des pathogènes. Ces mécanismes adaptatifs sont présentés dans les parties suivantes.

### 1.2.2. Les mécanismes évolutifs d'émergence des pathogènes : les hypothèses écologiques

Une maladie est dite émergente lorsqu'elle est décrite sur un nouvel hôte, dans une nouvelle région ou suite à une augmentation de la pathogénicité (Anderson *et al.*, 2004). Quatre grands mécanismes évolutifs peuvent conduire à l'émergence des agents pathogènes dans les agrosystèmes (Stukenbrock & McDonald, 2008).

Le premier modèle d'émergence proposé est le transfert horizontal de gènes. Ce mécanisme peut impliquer le transfert d'un seul gène, de régions génomiques voir même des chromosomes entiers. Sous ce scénario, le transfert de gènes impliqués dans la pathogénicité est effectué entre deux espèces reproductivement isolées (Friesen *et al.*, 2006). Ainsi, un organisme non pathogène sur une espèce hôte peut le devenir suite à un transfert de gène. Ainsi, l'émergence de *Pyrenophora tritici-repentis* sur blé serait dû à un transfert horizontal d'un gène codant une toxine, la Toxa (toxine essentielle à la virulence) provenant de *Phaeosphaeria nodorum* (Friesen *et al.*, 2006). Moran et Jarvik (2010) exposent un cas étonnant de transfert entre un puceron vert et un champignon. Le transfert horizontal de gènes de la voie de la biosynthèse des caroténoïdes du champignon vers le puceron vert, permet à ce dernier de synthétiser du torulène. Les pucerons rouges ayant subi ce transfert de gènes échapperaient mieux à la prédation que les pucerons verts.

Le deuxième modèle d'émergence est le saut d'hôte où le pathogène colonise une nouvelle gamme d'hôte jusqu'alors non infectée. Ce scénario évolutif est favorisé par les activités humaines qui dispersent et propagent les agents pathogènes hors de leur zone d'origine. Une étude de Zaffarano (2008) décrit le passage de *Rhynchosporium secalis*, de graminées sauvages au riz et à l'orge. *Magnaporthe oryzae* aurait « sauté » du millet *Setaria* vers le riz à partir duquel il aurait infecté les adventices communément associées à la riziculture : *Leersia hexandra* et *Panicum repens* (Couch *et al.*, 2005). Dans une revue récente, Giraud *et al.* (2010) présente ce modèle comme étant le plus courant des cas d'émergence de nouvelles maladies fongiques.

Selon le troisième modèle d'émergence, la co-différenciation (ou host-tracking) consiste en l'évolution d'un agent pathogène lors du processus de domestication de son hôte (Stukenbrock & McDonald, 2008). La forte sélection des traits agronomiques favorables chez les hôtes « sélectionnerait » parallèlement les pathogènes les mieux adaptés à ces génotypes d'hôtes. Dans ce modèle, l'agent pathogène et l'hôte partageraient le même centre d'origine et

subiraient des pressions d'évolution réciproques (Janzen, 1980). La littérature décrit plusieurs exemples de pathogènes qui auraient suivi ce scénario : *Ustilago maydis* sur le maïs (Munkacsí *et al.*, 2008), *Phytophthora infestans* sur pomme de terre (Gomez-Alpizar *et al.*, 2007), *Mycosphaerella graminicola* sur blé (Stukenbrock *et al.*, 2007 ; Torriani *et al.*, 2011) et *Venturia inaequalis* sur le pommier (Gladieux *et al.*, 2010 ; Lê Van *et al.* 2012).

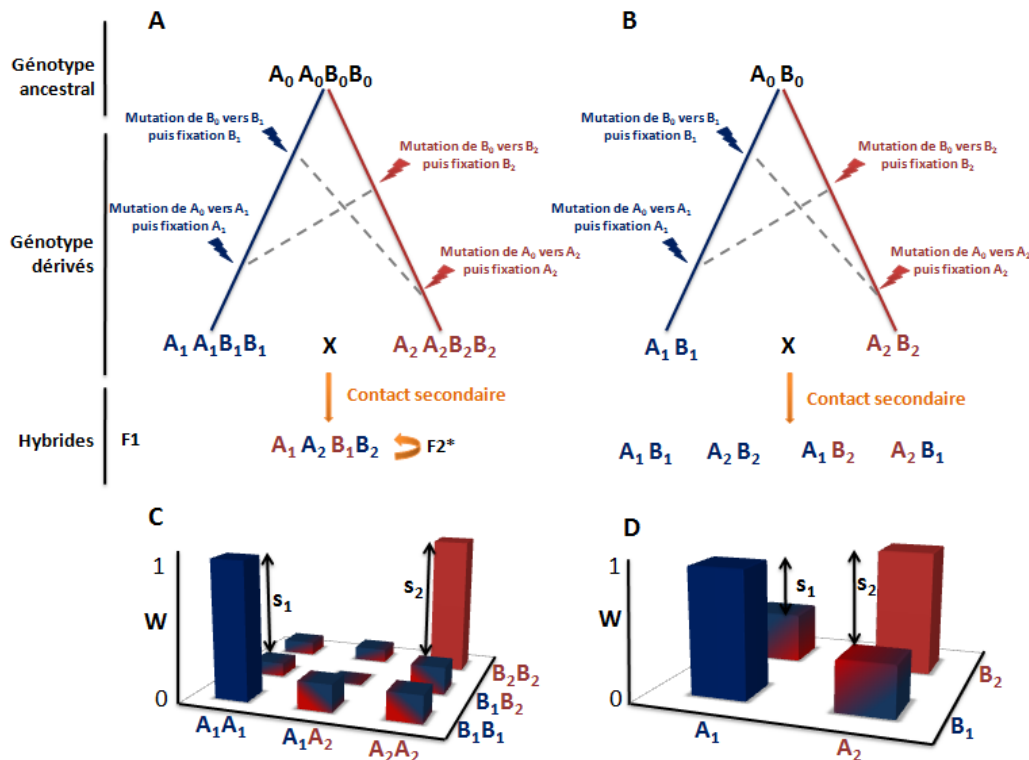
Le dernier modèle majeur d'émergence est l'hybridation interspécifique pouvant parfois être à l'origine de l'émergence d'hybrides capables d'infecter un nouvel hôte ou bien de combiner des gammes d'hôtes des parents (Ioos *et al.*, 2006 ; Brasier 1995). Ainsi, le croisement entre *Ophiostoma ulmi* et *Ophiostoma novo-ulmi* a permis l'émergence d'un hybride capable lui aussi d'infecter les ormes (Bates *et al.*, 1993). Récemment, ce modèle a été invoqué pour décrire l'émergence de *Mycosphaerella graminicola* à partir de l'hybridation de deux espèces sauvages *Zymoseptoria ardabilis* et *Z. pseudotritici* (Stukenbrock *et al.*, 2012a, b).

Ces différents modèles sont avant tout basés sur l'adaptation à l'hôte, et supposent donc que la divergence entre populations vivant sur des hôtes différents soit principalement d'origine écologique. Pourtant d'autres hypothèses moins liées à l'origine de l'hôte peuvent expliquer l'émergence de l'isolement reproductif dans les populations de pathogènes de plantes.

### **1.2.3. Hypothèses non écologiques : les incompatibilités Dobzhansky-Muller**

Alors que les modèles de divergence écologique cités précédemment ont pour principale conséquence l'isolement pré-zygotique, d'autres hypothèses impliquent davantage l'isolement post-zygotique lié à une baisse de la valeur adaptative des hybrides entre populations divergentes. Un tel modèle suppose une divergence préalable à l'hybridation, suivie d'une remise en contact entre ces populations divergentes. La première phase de divergence permet l'accumulation d'adaptations différentielles dans chacune des deux populations, mais aussi d'incompatibilités génétiques de type Dobzhansky-Muller (Dobzhansky, 1937 ; Muller, 1942) aussi appelées DMI. Ce modèle permet d'appréhender la mise en place d'un isolement post-zygotique intrinsèque sans nécessité de franchir les vallées adaptatives ou que les mutations soient délétères (Orr, 1995). Ce modèle implique la fixation de plusieurs loci neutres ou avantageux dans leur fond génétique habituel mais délétères

lorsqu'ils sont mélangés à d'autres allèles dérivés provenant d'un autre fond génétique. Dans sa forme la plus simple, la dépression hybride est alors due à des interactions épistatiques négatives entre deux locus bi-alléliques (A et B) comme expliqué dans la Figure 5.



**Figure 5 : Evolution de l'isolement reproductif sous le modèle de Dobzhansky-Muller pour un organisme diploïde (A) ou haploïde (B).**

A partir d'un génotype ancestral, de nouveaux allèles sont créés par mutation et se fixent par dérive génétique ou par sélection naturelle. Les allèles fixés peuvent alors être incompatibles entre les deux fonds génétiques. Ces incompatibilités, ici représentées par les lignes pointillées, sont alors révélées en cas de contact secondaire entre les deux populations. Les allèles d'incompatibilités sont ici considérés dominants et présentent des effets non-additifs. Les figures C et D représentent les valeurs sélectives associées aux génotypes de la F2 chez le diploïde et la F1 chez un haploïde et  $s_1$  et  $s_2$  représentent l'intensité de la sélection contre les hybrides. Notons que la distorsion de ségrégation chez un haploïde n'est observable que si  $s_1 \neq s_2$ . L'astérisque indique que la F2 n'est possible que si la génération F1 est viable, la production d'une F2 est donc rare si les allèles dérivés sont dominants.

L'existence de telles incompatibilités chez les organismes eucaryotes reste encore limitée aux organismes modèles tels que *Drosophila* (Sawamura & Yamamoto, 1997 ; Presgraves *et al.*, 2003; Brideau *et al.*, 2006 ; Ferree & Barbash, 2009 ; Phadnis & Orr, 2009; Tang & Presgraves, 2009), *Arabidopsis thaliana* (Bomblies *et al.*, 2007), *Saccharomyces* (Lee *et al.*, 2008 ; Chou *et al.*, 2010), *Mus musculus* (Mihola *et al.*, 2009). Si les DMI semblent *a priori* impliquer des mutations non-synonymes quelques études ont pu montrer que ils pouvaient aussi impliquer des régions non-codantes (Barbash *et al.*, 2003 ; Brideau *et al.*, 2006 ; Maheshwari & Barbash, 2011). A l'heure actuelle il n'existe pas d'étude permettant d'indiquer quelles classes de gènes seraient préférentiellement impliquées dans les DMI, ni même s'il y en aurait.

La présence des barrières post-zygotiques non-écologiques est donc particulièrement liée à l'existence d'une histoire évolutive particulière : celle des contacts secondaires. Pour autant la vraisemblance d'une telle situation chez les champignons phytopathogènes reste encore à démontrer.

#### **1.2.4. Les contacts secondaires chez les agents phytopathogènes**

L'hybridation entre taxons de pathogènes fongiques est considérée comme cause d'émergence de nouvelles maladies (Stukenbrock & McDonald, 2008). Cependant, les différents exemples impliquent des espèces différentes pouvant co-infecter les mêmes hôtes ou se retrouver ensemble à l'état saprophytique (Morris *et al.*, 2009). Qu'en est-il de l'hybridation intraspécifique de populations divergentes mais vivant sur des gammes d'hôtes similaires ? Cette hypothèse du contact secondaire entre populations divergentes, bien que courante chez d'autres organismes eucaryotes (Bierne *et al.*, 2011) est souvent négligée dans la littérature sur les pathogènes, privilégiant de loin les théories écologiques. Néanmoins, chez les organismes pathogènes de plantes, les probabilités d'observer une remise en contact entre deux populations divergentes sont loin d'être négligeables étant donné les opportunités qu'offrent les activités humaines. En effet, les pathogènes fongiques, en particulier les ascomycètes, possèdent généralement des spores peu favorables à des trajets à longues distances (McDonald et Linde, 2002), ce qui autorise l'isolement génétique de populations géographiquement distantes. Les remises en contacts sont surtout assurées par les activités humaines telles que le commerce mondial, l'importation de nouvelles cultures ou encore

l'exploitation de ressources génétiques d'origine sauvage. En effet, la redistribution des plantes d'origine anthropique peut autoriser la remise en contact secondaire des populations de pathogènes vivant sur ces hôtes, de manière beaucoup plus fréquente que le manque d'étude sur le sujet laisse à croire. C'est lors de ces événements que les différentes barrières écologiques et non-écologiques sont évaluées respectivement dans différents habitats et différents fonds génomiques (Kruuk *et al.*, 1999 ; Bierne *et al.*, 2011). C'est donc dans ces zones de remise en contact ou zones d'hybridation que l'on peut étudier la proportion de l'implication de l'ensemble des barrières, endogènes et exogènes, à l'isolement reproducteur. On comprend mieux pourquoi ces zones d'hybridation sont communément qualifiées de "laboratoires naturels de la spéciation" (Hewitt, 1988). L'étude de l'histoire évolutive des populations présentant des phénotypes contrastés, tels que nous pouvons les attendre dans le cas de la transition entre milieu sauvage et anthropisé, semble donc indispensable à l'interprétation des barrières d'isolement reproductif. Ainsi il n'est pas étonnant que les hypothèses *a priori* adaptatives n'interprètent les barrières que sous le seul angle écologique.

### **1.3. Diversification post-domestication : conséquences évolutives vis à vis de l'agent phytopathogène**

Le passage de l'environnement sauvage aux agroécosystèmes impacte donc potentiellement l'évolution des plantes hôtes sur un certain nombre de caractéristiques de leurs traits d'histoire de vie (cf § 1.1.2). Nous pouvons énumérer les potentielles conséquences de la domestication et de l'intensification de l'agriculture sur l'évolution des pathogènes présents sur les hôtes, animaux et végétaux. De manière synthétique, l'évolution de l'agriculture chez l'hôte a pour conséquence l'augmentation des populations génétiquement identiques. Les agroécosystèmes représentent un environnement plus favorable aux populations de pathogène adaptées à ces hôtes (Stukenbrock & McDonald, 2008). Hormis les études portant sur les espèces fongiques déjà domestiquées, peu d'études traitent de l'impact de la domestication chez les organismes fongiques (*Saccharomyces cerevisiae*: Gu *et al.*, 2005 ; *Aspergillus oryzae*: Rokas, 2009). Cependant, ces espèces ne sont pas pathogènes mais utilisées pour leur capacité de fermentation. A notre connaissance, l'impact de la domestication de l'hôte sur leur pathogène associé a été étudiée chez deux espèce : *Mycosphaerella graminicola* pathogène du blé



(Stukenbrock *et al.*, 2011) et *Magnaporthe oryzae* (Saleh *et al.*, 2014). Dans ces deux études, une conséquence observée chez les pathogènes lors de la domestication de leur hôte est la diminution de la diversité génétique chez le pathogène. Cette tendance est aussi observable chez le nématode *Trichinella spiralis* pathogène du porc domestiqué. On observe, en effet, une forte différenciation génétique entre les isolats provenant de la zone ancestrale de domestication du porc en Asie et les isolats provenant des élevages (Rosenthal, 2008). Les populations de nématode présentes sur porc domestique présentent une forte diminution de diversité par rapport aux populations sauvages.

A ce processus de domestication s'est ajoutée plus récemment l'intensification de l'agriculture et des échanges commerciaux qui influencent considérablement la structure et la dynamique des populations de pathogènes (Couch *et al.*, 2005 ; Stukenbrock *et al.*, 2007). L'intensification de l'agriculture a augmenté la taille des mêmes populations de pathogènes, comme chez le champignon pathogène du blé *M. graminicola* (Stukenbrock *et al.*, 2007). Or cette augmentation de la taille a des effets sur la diversité génétique abordée précédemment. Elle permet de limiter la dérive génétique et par conséquent d'augmenter la diversité chez l'espèce pathogène à l'échelle mondiale (Gladieux *et al.*, 2010). Les échanges commerciaux permettent quant à eux de maintenir la variabilité chez les populations de pathogènes, tout comme chez les hôtes (Salvaudon *et al.*, 2008).

De plus, l'intensification de l'agriculture peut aussi modifier les cycles de reproduction chez les espèces présentant des cycles sexués et asexués. En effet, dans un milieu favorable au développement de l'agent pathogène comme les agroécosystèmes, la dissémination asexuée est favorisée par le processus de domestication contrairement à la reproduction sexuée qui est généralement plus longue. (Stukenbrock & McDonald, 2008). Il en découle alors une augmentation de la taille de la population de pathogènes sur plantes cultivées.

Enfin, dans les agrosystèmes, les pathogènes profitent de la réduction de diversité des hôtes et de la disponibilité en éléments nutritifs pour augmenter leur dissémination. La transmission de maladies s'effectue plus facilement du fait de la configuration des agrosystèmes (Stukenbrock & McDonald, 2008). En effet, la plus forte densité d'hôtes génétiquement plus homogènes favorise la dissémination des souches adaptées. On peut alors observer de fortes expansions démographiques de souches virulentes dans la population du pathogène (Thrall & Burdon, 1999).

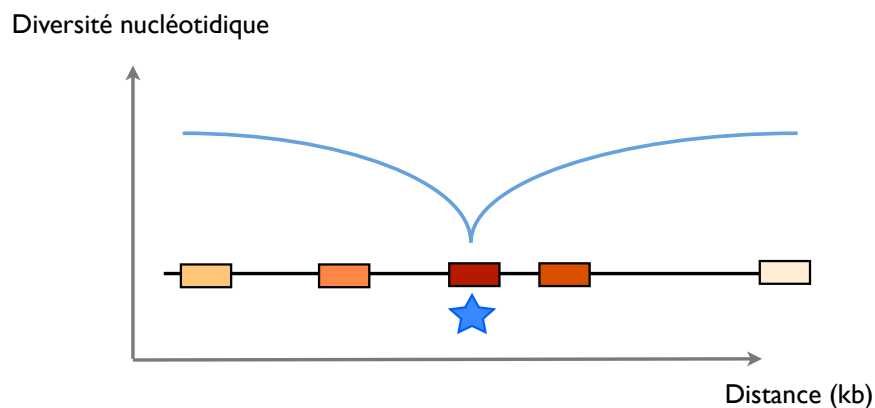
Il apparaît donc que la transition vers les agrosystèmes des champignons phytopathogènes s'est donc traduite par une évolution des traits d'histoire de vie. Comme nous

l'avons vu, ces évolutions peuvent concerner les modes de reproduction et le succès de transmission à un hôte voisin. Toutefois, n'existe-t-il pas d'autres traits d'histoire de vie comme la dispersion ? En effet, on peut supposer que la configuration particulière des agroécosystèmes en terme de densité et d'homogénéité d'hôte a pu affecter le potentiel de dispersion des pathogènes. Ne peut-on pas imaginer que les paramètres du succès de dispersion en milieu sauvage, où la densité d'hôte serait plus faible et l'hétérogénéité génétique plus importante, aient pu évoluer vers un autre optimum adaptatif en milieu agricole ?

L'étude de l'impact des changements environnementaux liés à l'évolution de l'agriculture chez les pathogènes nous offre une formidable opportunité d'étudier un processus évolutif sur un temps court. L'apparition de l'agriculture date de 12 000 ans sur le continent eurasien, ce qui peut être considéré comme une époque récente d'un point de vue évolutif. Toutefois, la domestication n'est pas le seul processus où des souches sauvages de pathogènes sont confrontées aux agroécosystèmes. En effet, l'utilisation de ressources génétiques sauvages chez les plantes en vue de l'introgession de gènes d'intérêt agronomique dont la résistance à certaines souches de pathogènes, constitue une excellente opportunité pour des souches sauvages à infecter de nouveaux hôtes domestiques (Leroy *et al.*, 2013). Travailler sur des temps évolutifs courts présente plusieurs avantages. La divergence récente entre les populations de pathogènes sauvages et de cultures doit faciliter la distinction entre dérive et sélection. En effet, les changements environnementaux comparables à ceux rencontrés lors des transitions entre milieu sauvage et anthropisé sont censés privilégier les événements de sélection diversifiante (Stukenbrock & McDonald, 2008). Ce régime favorise la fixation alternative d'allèles favorables dans chaque milieu. Cependant, les études de recherche de locus sous sélection sont souvent complexes à cause des fixations d'allèles dues à la dérive plus qu'à la sélection. Toutefois le processus de fixation alternative d'allèles neutres est un processus long, de l'ordre de 4 fois la taille efficace de la population, supérieur au temps de fixation d'une mutation avantageuse. Ainsi sur des temps courts, la proportion de locus différenciellement fixés à cause de la sélection est-elle plus importante que pour des temps plus longs ? De plus, comme la fixation d'une mutation avantageuse est rapide, celle-ci n'a pas le temps de recombiner avec les polymorphismes voisins qui ségrègent dans la population. On observe alors une forte réduction du polymorphisme au site même de la mutation favorable mais aussi à son voisinage, dont la longueur dépend de l'intensité du coefficient de sélection. Ce processus est appelé balayage sélectif (Maynard-Smith & Haigh, 1974 ; Encadré I) et s'estompe avec le temps, la recombinaison favorisant une augmentation du polymorphisme au

### Encadré I : Le balayage sélectif

Le balayage sélectif (selective sweep) considère l'effet d'une mutation avantageuse à un locus donné, sur le polymorphisme des locus neutres qui sont liés (Maynard-Smith et Haigh, 1974). Lorsqu'une mutation avantageuse apparaît dans un génome et se fixe au cours du temps, la région où se localise cette mutation se retrouve dépourvue de mutations neutres qui s'accumulent généralement aux sites silencieux. A plus longue distance de cette mutation, l'effet du balayage s'estompe progressivement, jusqu'à ne plus être repérable (Figure 6). Le polymorphisme neutre peut être juste réduit (et non perdu) si la mutation avantageuse se localise dans une région où il y a de la recombinaison. Ce modèle prédit (i) une corrélation positive entre le taux de recombinaison local et le polymorphisme neutre (si les mutations se fixent souvent dans tous les génomes et si le taux de mutation n'est pas trop fort) (Payseur & Nachman, 2000) (ii) mais également la fixation de mutations délétères liées aux mutations avantageuses.



**Figure 6. Diversité nucléotidique lors d'un balayage sélectif**

La diversité nucléotidique décroît là où la sélection a amené la fixation d'un nouvel allèle avantageux (étoile bleue). Cette décroissance de la variabilité permet d'identifier les régions touchées par la sélection naturelle

voisinage de la mutation favorable. Encore une fois, il est plus simple de travailler sur des événements évolutifs récents afin de détecter plus facilement les processus sélectifs.

*A contrario*, l'identification précise des zones sous sélections sera rendue plus difficile. En effet, les zones impliquées dans le balayage sélectif étant plus larges, la variance des fréquences alléliques entre populations telle que mesurée par le  $F_{st}$ , couvrira une zone génomique plus importante. En d'autres termes, la mutation avantageuse sera génétiquement liée à une zone neutre plus large. L'utilisation de scans génomiques basés sur les  $F_{st}$  (Lewontin & Krakauer, 1973 ; Beaumont & Nichols, 1996 ; Vitalis *et al.*, 2001 ; Beaumont & Balding, 2004 ; Foll & Gaggiotti, 2008) identifiera non seulement les polymorphismes sous sélection mais aussi les polymorphismes neutres qui leur sont liés.

L'interprétation des polymorphismes détectés par les méthodes de scans génomiques nécessite pour autant une bonne connaissance de l'histoire évolutive des populations étudiées. En effet, comme le montrent Bierne *et al.* (2011), la méconnaissance de l'histoire évolutive, et donc de l'histoire démographique des populations est une importante source de biais dans l'interprétation des loci détectés par ces tests de neutralité. Bierne *et al.* (2011) insistent en particulier sur les cas de contacts secondaires rarement testés dans les scénarios évolutifs et qui pourtant pourraient être plus fréquents que supposés chez les pathogènes. L'inférence démographique est donc un préalable indispensable à la détection de polymorphismes adaptatifs.

Les dernières années ont connu une innovation importante des méthodes d'inférence démographique. Nous détaillerons principalement deux de ces méthodes : la méthode ABC (Approximate Bayesian Computation, Beaumont *et al.*, 2002) et la méthode de diffusion implémentée dans le logiciel *δaδi* (Gutenkunst *et al.*, 2009). L'utilisation de marqueurs génétiques (séquences, SNP) et des méthodes d'inférence de scénarios démographiques permettent de retracer l'histoire démographique de populations ayant divergées. Actuellement, les programmes les plus utilisés sont basés sur des inférences bayésiennes et des simulations des chaînes de Markov Monte Carlo (MCMC). Or, les paramètres essentiels pour étudier le processus de divergence avec flux de gènes ne sont pas inférés avec ces méthodes (taux de migration, temps de divergence). Ces modèles basés sur la théorie de la coalescence ne permettent pas d'estimer ces paramètres simultanément (Nielsen & Wakeley, 2001). D'autres logiciels comme IMA2 permettent d'inférer les paramètres démographiques associés à la divergence. Cependant, l'utilisation de ces logiciels reste sommaire face au calcul de la vraisemblance de scénarios évolutifs trop complexes (Csilléry *et al.*, 2010). Ainsi une nouvelle technique, l'ABC, permet de comparer des scénarios évolutifs beaucoup plus

complexes et réalistes (Encadré II). Grâce à cette approche, la comparaison de modèles avec ou sans flux de gènes est possible afin de mieux aborder le processus de diversification. L'approche ABC se résume en 3 étapes : (i) une étape de simulation de jeux de données et de calcul de statistiques résumées, (ii) une étape de sélection des simulations, (iii) une étape d'estimation des distributions *a posteriori* des paramètres par régression linéaire (Encadré I et Figure 7).

### Encadré II : Approximate bayesian computation (ABC)

Les populations naturelles ont des histoires démographiques complexes. Au cours de ces dernières années, les approches se basaient souvent sur la technique de Markov Chain Monte Carlo (MCMC). Cependant leur portée était limitée à l'utilisation de scénarios évolutifs simples de par la difficulté à calculer leur vraisemblance. Afin de contourner cette estimation de la vraisemblance, une nouvelle approche appelée Approximate Bayesian Computation (ABC) basée sur des statistiques résumées a vu le jour (Beaumont *et al.*, 2002). Cette méthode permet de simuler des millions de généalogies avec différentes valeurs de paramètres sous

différents modèles avec ou sans flux de gènes. Les distributions postérieures sont calculées à partir des simulations dont les valeurs de statistiques résumées sont les plus proches de celles du jeu de données réelles (Figure 7).

#### • Modèles

Avant de traiter les données, les scénarios démographiques sont proposés suivant différentes hypothèses comprenant la taille des populations, les temps de divergence. Lors de l'élaboration des modèles, des connaissances préliminaires sur le modèle biologique et des données historiques sont nécessaires. La conception des modèles doit cependant être limitée par leur nombre et leur complexité afin de trouver l'explication la plus parcimonieuse.

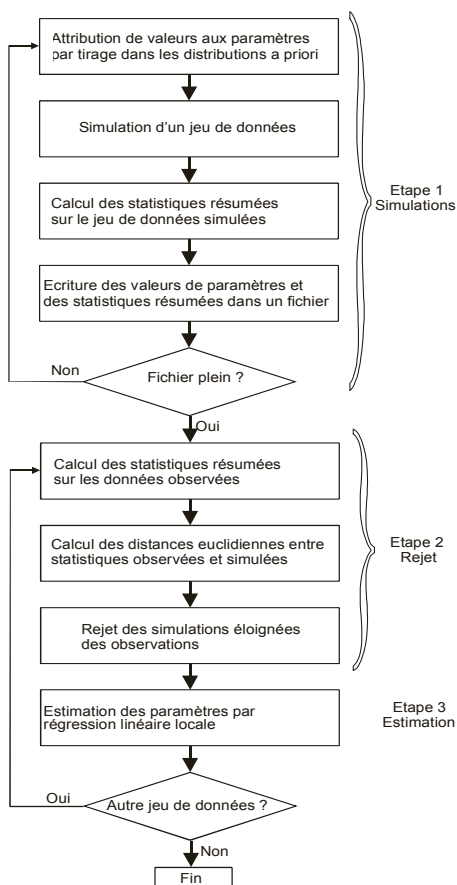


Figure 7. Principales étapes de l'ABC

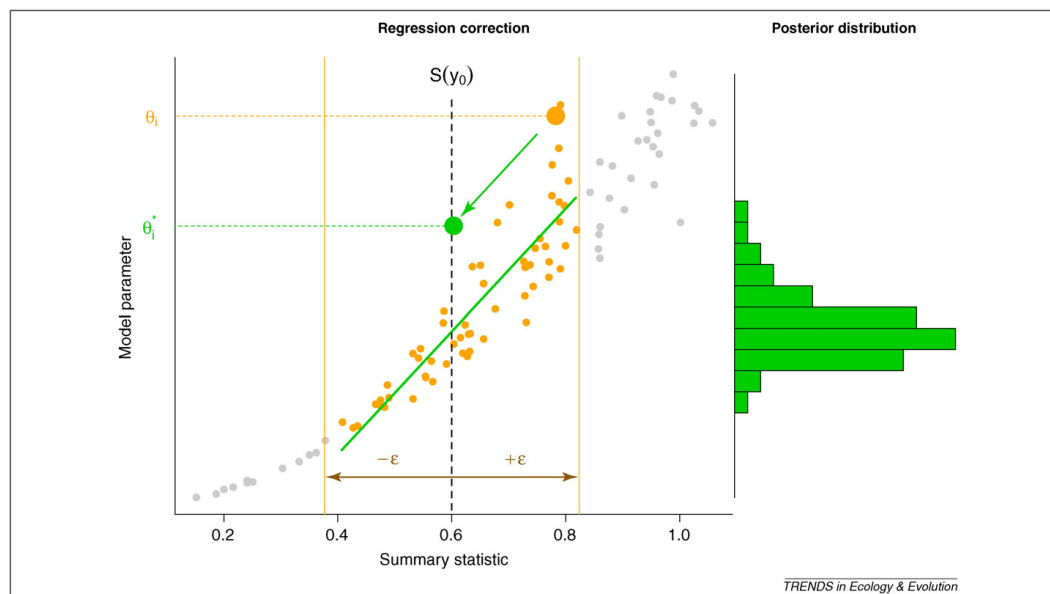
*Encadré II : suite*

- Inférence et choix du modèle

L'étape suivant l'écriture des différents scénarios est la simulation des données. Sous chaque modèle, la valeur  $x$  est échantillonnée plusieurs fois afin de simuler un jeu de données  $y_i$  (Figure 8). Il s'agit ensuite de comparer de la valeur statistique résumé associée à  $S(y_i)$  à la valeur de la statistique du jeu de données observée  $S(y_0)$ . Si  $S(y_i) - S(y_0) < \text{seuil de tolérance } \varepsilon$ , alors la valeur de paramètre est acceptée (Points jaunes, Figure 8). Ces valeurs sont ensuite ajustées par régression linéaire (Leuenberger & Wegmann, 2009) (ligne verte, Figure 8), pour enfin donner la nouvelle valeur de paramètre (Point vert, Figure 8). Sous chaque modèle, ces différentes étapes sont réalisées, puis les modèles sont comparés entre eux par l'estimation des probabilités postérieures.

- Fiabilité

La fiabilité de la procédure s'observe via la concordance entre les courbes de rejet et la distribution postérieure des paramètres. Lorsque les deux possèdent la même distribution, les estimations sont considérées comme fiables. Sinon, l'estimation des paramètres est biaisée, traduisant un problème d'ajustement du modèle aux données.



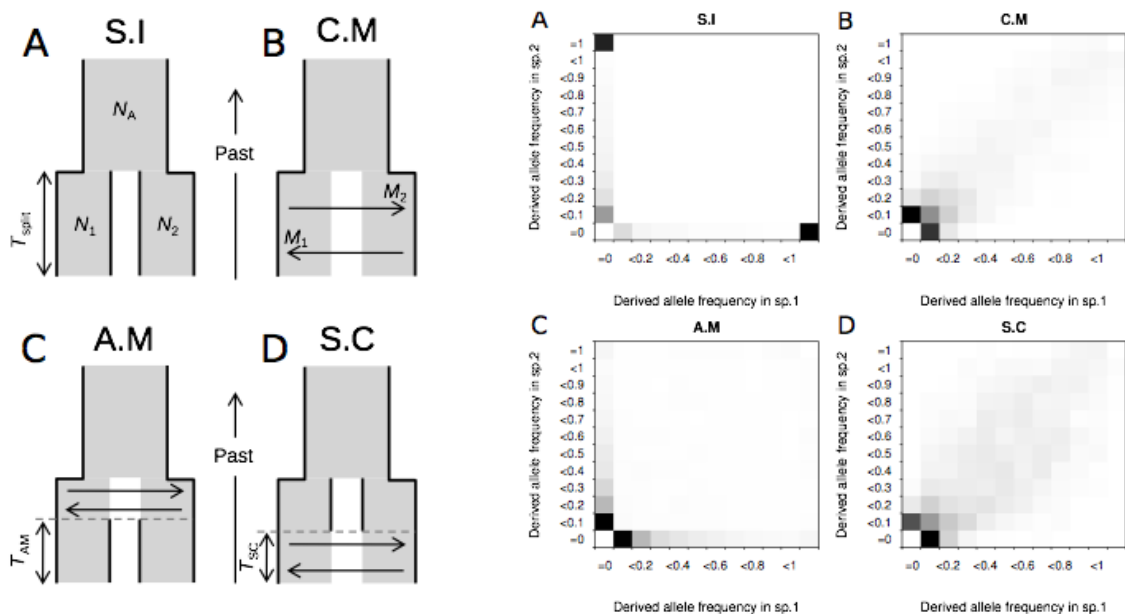
**Figure 8. Ajustement par régression linéaire lors d'une ABC (Csilléry *et al.*, 2010)**

Le paramètre  $\varepsilon$  mesure la taille de la fenêtre au sein de laquelle les simulations sont acceptées. Les paramètres  $\theta_i$  représentent les valeurs des paramètres et les  $\theta_i^*$  correspondent aux valeurs des paramètres après ajustement.

L'analyse par la méthode ABC permet d'établir le cadre évolutif le plus probable pour la mise en place des barrières aux flux de gènes (Roux, 2010). Cette analyse peut prendre en compte deux types de barrières : sous un modèle homogène ou hétérogène, qui correspondent aux différentes barrières possibles tout le long du génome. L'intégration de ces deux types de migration est essentielle afin de ne pas passer à côté du bon modèle démographique.

L'inférence du scénario évolutif peut également s'effectuer en utilisant la méthode de diffusion implémentée dans *δaδi* qui permet d'obtenir des spectres de fréquences joints (Gutenkunst *et al.*, 2009) (Figure 9). Les spectres de fréquences joints représentent la fréquence de l'allèle dérivé (en référence à un groupe externe) dans chaque population fille. On peut alors quantifier la part de sites polymorphes en commun entre les 2 populations divergentes et les sites fixés dans chaque population.

Les différences sont fixées lorsque la fréquence de l'allèle dérivé est égale à 1 dans une espèce et est égale à 0 dans l'autre. Pour les sites polymorphes partagés entre les 2 espèces, la fréquence est supérieure à  $> 0$  dans une espèce et inférieure à 1 dans l'autre. Ainsi pour un scénario, un spectre de fréquence joint est attribué.



**Figure 9. Exemple de scénarios évolutifs (à gauche) et les spectres de fréquences associés (à droite)**

$M_1$  et  $M_2$  : taux de migration     $a$  : population ancestrale    1 et 2 : population 1 et 2     $N_x$  : taille efficace de la population  $x$  exprimée en nombre d'individus     $T_{split}$  : Temps de divergence avec la population ancestrale     $T_{SC}$  : Temps de remise en contact secondaire     $T_{ISO}$  : Temps d'isolement     $T_{AM}$  : Temps rupture migration

Chaque case noire représente la case la plus représentée. L'intensité de la coloration des autres cases dépend de leurs importances relatives à la classe la plus représentée

## 2. Le pathosystème *Venturia inaequalis* / pommier

### 2.1. La tavelure du pommier

#### 2.1.1. Importance économique

*Venturia inaequalis* est un champignon ascomycète haploïde et hémibiotrophe, responsable de la tavelure du pommier. D'autres hôtes moins importants économiquement appartenant aux genres *Sorbus*, *Cotoneaster*, *Crataegus*, *Heteromeles*, *Kageneckia*, *Pyracantha* et *Eriobotrya* ont également été décrits (Menon, 1956 ; Raabe & Gardner, 1972 ; Sivanesan, 1977 ; Le Cam *et al.*, 2002). Elle est présente dans toutes les zones de production du pommier. Cette maladie est particulièrement dommageable au printemps dans les régions humides et fraîches.

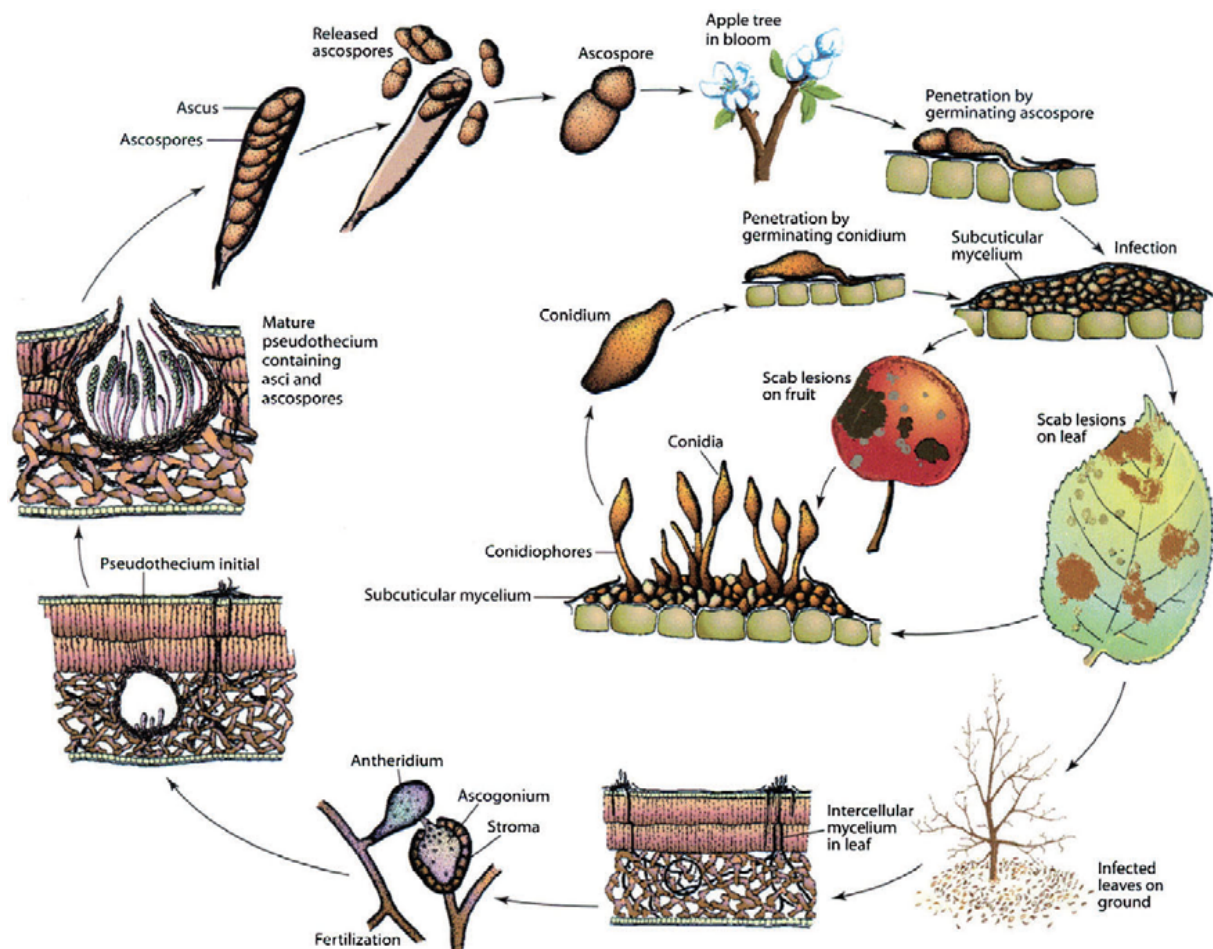
La tavelure est la maladie du pommier la plus préjudiciable économiquement. Même si elle ne menace pas la vie de son hôte, les pertes économiques liées à la tavelure résultent principalement des infections des fruits, détériorant leurs aspects visuels et les rendant impropres à la commercialisation et à l'exportation. Sachant que plus de 90% de la production mondiale repose sur des variétés sensibles, l'enjeu de maîtriser le développement de la tavelure et des autres maladies infectant le pommier est primordial.

La lutte contre la tavelure nécessite une combinaison entre un apport fongicide et des pratiques culturales raisonnées (Holb, 2006). Afin de réduire le nombre de traitements fongicides, des modèles de risques d'infection en fonction des conditions climatiques ont été développés et permettent d'effectuer les traitements fongicides aux meilleurs moments de l'infection (Mills & Laplante, 1951 ; Carisse *et al.*, 2009). De plus, dans le contexte de réduction de traitements chimiques, d'autres méthodes de lutttes sont employées contre la tavelure telles que la prophylaxie, la modification des méthodes culturales, la lutte biologique (Hossain *et al.*, 2009) et la résistance variétale. Les coûts liés aux différents moyens de lutte contre cette maladie restent conséquents, même s'ils restent nettement moindres que les coûts engendrés par la main d'œuvre nécessaire tout au long de l'année sur une culture de pommier.



### 2.1.2. Symptomatologie et cycle de vie

Son cycle de vie comprend deux phases : (i) une phase parasitaire au cours de laquelle se déroulent plusieurs cycles de reproduction asexuée et (ii) une phase saprophytique, sur la litière foliaire, au cours de laquelle a lieu la reproduction sexuée et qui permet la survie hivernale du champignon sous forme de périthèces (Figure 10).



**Figure 10. Cycle biologique de *Venturia inaequalis* (Agris, 2005)**

Le cycle de vie de *Venturia inaequalis* comprend une phase sexuée parasitaire et de multiples phases de reproduction asexuée.

L'infection débute au printemps lorsque les jeunes feuilles sont les plus sensibles et que les conditions de température et d'humidité déclenchent la projection d'ascospores à partir des périthèces. Au niveau de la litière foliaire, la projection des ascospores lors

d'évènements pluvieux s'effectue sur de très faibles distances dans un premier temps (quelques millimètres). Puis elles sont dispersées sur de plus grandes distances par le vent, de l'ordre d'une dizaine de mètres (Aylor, 1998 ; McHardy *et al.*, 2001 ; Leroy *et al.*, 2013).

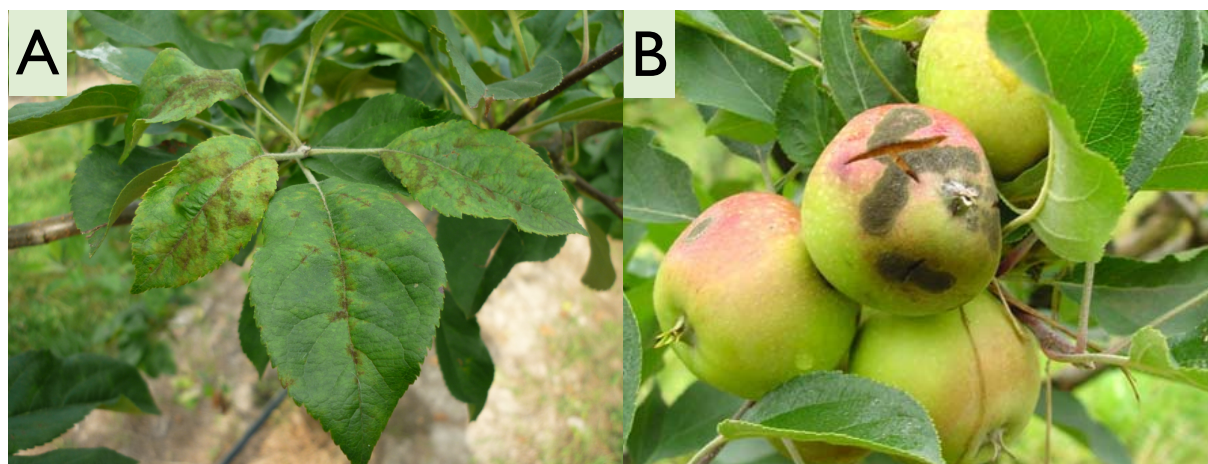
La germination des ascospores se fait en présence d'eau libre sur la surface des feuilles. Le tube germinatif va différencier un appressorium qui constitue l'organe de pénétration (Smereka *et al.*, 1987 ; Jha *et al.*, 2009). Sous l'appressorium, le mycélium perfore la cuticule. Après pénétration, le champignon développe alors une structure pseudo-parenchymateuse, nommée stroma, à partir de laquelle un mycélium sous-cuticulaire se différencie puis des conidiophores sont abondamment produits. Les conidiophores, portant à leur extrémité les conidiospores, vont déchirer l'épiderme afin de libérer les conidiospores qui constituent l'inoculum secondaire. Elles sont alors dispersées par le vent ou la pluie sur de courtes distances (quelques mètres). Les capacités intrinsèques de migration des ascospores et conidies étant relativement faibles, la migration du champignon à longue distance se fait principalement par le transport de matériel infecté.

Le développement du champignon au cours de ces deux phases est très dépendant des conditions climatiques. En effet dans les régions à hiver peu rigoureux, le pommier ne perd pas toutes ses feuilles et peut ainsi assurer une partie de sa survie hivernale selon le mode clonal.

Les symptômes causés par *Venturia inaequalis* sont visibles sur différents organes de la plantes : feuilles, fruits, pétioles, sépales, fleurs et bourgeons. Sur les feuilles, la tavelure produit des lésions pouvant être soit isolées, dispersées sur tout le limbe ou bien recouvrir la surface entière (Figure 11 - A). Généralement, les lésions se développent plus rapidement sur les jeunes feuilles à cause de l'existence de résistances ontogéniques présentes chez les feuilles plus âgées. L'infection se produit sur la face supérieure des feuilles en début et milieu de saison et sur la face inférieure en fin de saison. Les lésions sont d'abord de petites tailles, circulaires et translucides. Elles vont s'agrandir pour devenir de couleur brun foncé, et prendre un aspect velouté de par la production en masse de conidies. Durant le vieillissement des lésions, les feuilles vont se dessécher, se déformer, ce qui provoque une déchirure du limbe entraînant une défoliation plus ou moins importante en fonction de la gravité de la maladie.

Les pommes jeunes sont plus sensibles à l'infection de *Venturia inaequalis* que les fruits mûrs. L'infection provoque des déformations, des tâches et peut entraîner la chute des fruits lorsque le pédicelle est touché (Figure 11 - B). Quand l'infection est plus tardive, les lésions

se développent beaucoup plus lentement. Les lésions ne sont alors visibles à l'œil nu que lors de la sortie des fruits des salles de conservation après plusieurs mois de stockage.



**Figure 11. Symptômes de tavelure sur feuilles (A) et sur fruits (B)**

## **2.2. Processus de domestication et histoire évolutive**

### **2.2.1. Origine et domestication du pommier cultivé**

L'hôte le plus étudié de *Venturia inaequalis* est le pommier domestique (*Malus x domestica*). Le genre *Malus* serait apparu en Chine vers la fin de la période du Tertiaire, à l'époque Eocène, il y a environ 30 millions d'années (Phipps *et al.*, 1991). On décompte entre 8 et 79 espèces reconnues selon les auteurs (Juniper & Mabberley, 2006). Cette large variabilité est causée par la difficulté à différencier les espèces selon leurs caractéristiques morphologiques. De plus, cette différence du nombre d'espèces décrites résulte aussi de l'importance de l'hybridation et de la présence de polyploïdie chez le genre *Malus* (Potter *et al.*, 2007; Campbell *et al.*, 1991). L'utilisation de marqueurs moléculaires associée à des études phylogénétiques ont permis de préciser la taxonomie de certaines espèces de *Malus* (Robinson *et al.* 2001; Velasco *et al.*, 2010 ; Cornille *et al.*, 2012 ; Dunemann *et al.*, 2012). Cependant la plupart des études se sont focalisées sur la contribution de deux espèces sauvages *M. sylvestris* et *M. sieversii*, endémique de l'Europe et de l'Asie Centrale et respectivement.

Le pommier cultivé porte le nom de *Malus x domestica*, le « x » faisant référence au caractère hybride du pommier domestique (Korban & Skirvin, 1984). La ressemblance entre

les fruits des pommiers domestiques et ceux portés par *M. sieversii* (espèce endémique sauvage d'Asie Centrale) a mené rapidement Vavilov (1930) à proposer *M. sieversii* comme étant l'ancêtre ancestrale du pommier domestique. L'ensemble de la diversité morphologique observée chez cette espèce est présent aujourd'hui chez les pommiers domestiques, laissant envisager une sélection consciente des pommiers ancestraux (Juniper & Mabberley, 2006 ; Forsline *et al.*, 1994 ; Forsline, 1995). L'histoire de la domestication de la pomme est maintenant assez bien connue et documentée (Cornille *et al.*, 2014). Le pommier cultivé a effectivement été domestiqué en Asie centrale à partir de *Malus sieversii* (Velasco *et al.*, 2010 ; Cornille *et al.*, 2012.). *M. sieversii* est toujours présent sous forme de forêts entre les montagnes de Tien Shan (frontière du Kazakhstan, le Kirghizstan et Province du Xinjiang en Chine) et la mer Caspienne (Morgan & Richards, 1993). C'est au cours de la période néolithique et surtout pendant le développement de la Route de la soie que les pépins de pommiers ont été dispersés. Au cours des déplacements, les marchands et les animaux auraient disséminé les pépins marquant ainsi le début des cultures du pommier (Juniper & Mabberley, 2006 ; Wood, 2003 ; Harris *et al.*, 2002) . Le pommier domestiqué se serait ensuite propagé vers l'Ouest de l'Europe et vers l'Est jusqu'à la Chine (Juniper & Mabberley, 2006). La migration vers l'ouest a alors impliquée des contacts et hybridations avec les plus proches parents sauvages qui poussent le long de la Route de la Soie, *M. sylvestris* en Europe, *M. orientalis* dans le Caucase (Cornille *et al.*, 2013). De récents travaux ont ainsi montré que le pommier sauvage Européen *M. sylvestris* a largement contribué au pool domestique Européen (Cornille *et al.* 2012). Le pommier sauvage *M. sylvestris* présente une aire de répartition s'étendant de l'Europe de l'Est à l'Europe de l'Ouest. Le pommier aurait ensuite été introduit en Europe et en Afrique du Nord par les Grecs et les Romains il y a environ 2 000-2 500 ans (Forsline *et al.*, 2002 ; Juniper et Mabberley, 2006) et l'Europe aurait ensuite servi de tête de pont pour la dissémination du pommier à travers le monde après la découverte des nouveaux continents (Morgan & Richards, 1993)

### **2.2.2. Histoire évolutive des populations de *Venturia inaequalis* au sein du centre de domestication du pommier**

Le pathosystème *Malus spp.-Venturia inaequalis* est particulièrement pertinent pour étudier les changements des traits d'histoire de vie des agents pathogènes s'adaptant à des hôtes domestiques et des agrosystèmes. Sa reproduction se produit par croisement de souches de types sexuels opposés ayant infecté la même feuille. Cette reproduction exclusive entre

souches capables de se développer sur le même génotype faciliterait le processus d'isolement prézygotique et par conséquent une divergence écologique (Giraud *et al.*, 2010 ; Gladieux *et al.*, 2011). La capacité de dispersion des deux types de propagules produits par *V. inaequalis* (ascospores et conidiospores) est faible. La manière la plus favorable pour l'agent pathogène d'atteindre une dispersion à longue distance est liée aux activités humaines : le transport de fruits et de plantes infectées (McHardy, 1996 ; Holb *et al.*, 2004). Ces éléments supposent que la structure de population du pathogène est en lien avec l'histoire de la propagation de son hôte (Gladieux *et al.*, 2010).

#### a) *Structure des populations de Venturia inaequalis au Kazakhstan*

Les études génétiques antérieures sur *V. inaequalis* nous donnent beaucoup d'indices sur l'origine de l'agent pathogène. Une analyse réalisée à l'échelle mondiale de la diversité des populations *V. inaequalis* a d'abord montré que l'agent pathogène partageait une origine commune avec son hôte en Asie centrale (Gladieux *et al.*, 2008). L'impact de la domestication de la pomme sur la structure génétique de l'agent pathogène a alors été étudié (Gladieux *et al.*, 2010). Basé sur le polymorphisme de onze loci microsatellites, l'utilisation des méthodes d'assignations a permis d'identifier 3 populations : (1) une population Européenne regroupant des souches isolées à la fois sur le pommier sauvage endémique Européen, *M. sylvestris* et sur les pommiers domestiques en Europe; (2) une population Asiatique regroupant les souches isolées du pommier endémique *M. sieversii* présent en plaine dans les zones anthropisées d'Asie centrale (CAP : Central Asia Plains) ; (3) une population également isolée de *M. sieversii* mais restreinte géographiquement aux montagnes de l'Est du Kazakhstan (CAM : Central Asia Mountains). Cette dernière population est considérée comme étant une relique de la population ancestrale des populations actuellement présentes dans les écosystèmes agricoles du monde entier (Gladieux *et al.*, 2010).

#### b) *Estimation du temps de divergence*

Les probabilités postérieures du temps de divergence entre la population isolée des montagnes Kazakhes (CAM) et les 2 autres populations suggèrent une divergence évaluée entre 2 000 et 4 000 ans (Gladieux *et al.*, 2010). Cette estimation coïncide avec la période de domestication du pommier : 2 000 à 7 000 ans (Juniper & Mabberley, 2006). D'après les auteurs, le scénario le plus probable d'émergence serait donc un scénario de coévolution par host-tracking. L'agent pathogène aurait suivi son hôte lors de la domestication du pommier, lors de l'introduction des pommiers en Europe (Gladieux *et al.*, 2010). L'estimation du temps de



divergence présente un intervalle de confiance large (700 à 11 500 ans) qui est fortement dépendant du taux de mutations des locus microsatellites utilisés ainsi que de la variance dans le nombre de répétitions des microsatellites de la population ancestrale.

Sous ce scénario évolutif, les variants les plus favorables seront sélectionnés dans les agrosystèmes. Selon le modèle du balayage sélectif (Figure 6), on s'attend à observer une baisse de diversité aux loci impliqués dans l'adaptation aux conditions de culture des milieux anthropisés.

### c) *Evolution des traits d'histoire de vie de la tavelure*

Une étude portant sur les caractéristiques de pathogénicité de la population ancestrale (CAM) a montré que la domestication du pommier avait nettement modifié la pathogénicité de *V. inaequalis* (Lê Van *et al.*, 2012). A partir d'inoculations croisées de ces populations sur différents types d'hôtes (*M. sieversii*, *M. sylvestris* et *M. x domestica*), les auteurs ont démontré que les souches de la population CAM étaient incapables d'infecter les cultivars sensibles 'Gala' et 'Top Red Delicious' très déployés dans les vergers. Le processus de domestication se serait donc accompagné d'une acquisition de virulence et d'une augmentation de l'agressivité globale des populations de *V. inaequalis*, sous le scénario d'host-tracking entre le pommier et la tavelure (Lê Van *et al.*, 2012). Dans le cadre de ce projet, nous allons chercher à identifier d'autres traits d'histoire de vie qui auraient été modifiés lors du passage du compartiment sauvage vers le compartiment domestiqué (cf Chapitre 1). L'étude se base donc sur les populations de *V. inaequalis* du Kazakhstan considérées comme étant les descendantes des populations ancestrales. Les populations domestiques seront représentées par des échantillons prélevés au Kazakhstan et en Chine dans des zones anthropisées (CAP). Nous étudierons plusieurs comparaisons phénotypiques pour étudier ces changements de traits d'histoire de vie impliqués dans la dissémination de la tavelure, associés à la domestication de son hôte. L'impact de la domestication des plantes sur les organismes pathogènes est rarement étudié malgré son importance majeure pour comprendre les conséquences des activités humaines sur l'émergence des maladies.

Dans ce chapitre, nous discuterons de pestification, notion récemment abordée pour Saleh *et al.* (2014) qui définit les changements de traits d'histoire de vie chez les pathogènes suite à la domestication de leur hôte.

#### d) La résistance *Vf*

Le pommier sauvage *M. x floribunda* a fortement été utilisé dans les programmes d'amélioration car porteurs de gènes de résistances majeurs dont *Vf*. *M. x floribunda* serait un hybride interspécifique de deux espèces de pommiers sauvages utilisés pour l'ornementation. La découverte de la résistance par Hough (1944) a permis d'obtenir plus de 80 variétés commercialisées, faisant de la résistance *Vf*, la plus introgressée dans les cultivars de pommier (Crosby *et al.*, 1992 ; Gessler & Pertot, 2012). Pendant plus de 40 ans, l'absence de contournement a classé cette résistance *Vf* comme « durable » (Crosby *et al.*, 1992). Les premières observations d'infection des cultivars *Vf* ont été relevés sur des cultivars de *M. x domestica* et *M. x floribunda* en Allemagne en 1988 (Parisi *et al.*, 1993). C'est en 1995 dans un verger Normand, que la première mise en évidence de souches virulentes a été détecté en France (Parisi *et al.*, 2000). A l'heure actuelle, la résistance est largement contournée en Europe (Guérin & Le Cam, 2004). L'efficacité au champ est cependant préservée par la mise en place de mélanges variétaux et l'apport raisonné en fongicides. Des analyses génétiques réalisées dans les vergers cidricoles démontrent que les populations virulents partagent une origine commune et possède une diversité réduite comparé aux populations *non-Vf*. L'existence d'une forte différenciation entre les variétés *Vf* et non *Vf* et son maintien au cours du temps suggéraient un flux de gènes restreint entre les deux populations (Guérin *et al.*, 2007 ; Gladieux *et al.*, 2011). De plus, l'absence de deux marqueurs (lié au locus d'avulence *Vf*) dans les populations avirulentes suggérait que cette la population virulente n'a pas émergé des populations virulentes, mais résulterait d'une migration de souches à partir d'une source externe. Ainsi, les hypothèses de l'émergence de *Vf* se classent en 2 théories : (1) soit les populations virulentes sont apparues par mutation dans les agrosystèmes, (2) soit les populations virulentes préexistaient sur un hôtes alternatif à l'extérieur des agrosystèmes.

### 2.3. Questions de recherche

#### **La domestication du pommier a-t-elle provoqué des changements de traits d'histoire de vie chez *V. inaequalis* ?**

Dans cette partie nous testerons l'hypothèse d'un changement de traits d'histoire de vie chez *V. inaequalis* en comparant sur des caractères phénotypiques des souches sauvages et d'agroécosystèmes kazakhes. Comme évoqué plus haut, le passage de l'habitat sauvage aux vergers pourrait avoir un impact sur les paramètres de succès d'infection sur hôte compatible, et par là même, de dispersion efficace. Si d'éventuelles différences existent pour certains traits phénotypiques, nous testerons pour ces traits l'hypothèse d'une adaptation *de novo* aux milieux anthropisés par rapport à celle d'une sélection à partir d'une variation phénotypique existante dans les populations sauvages et qui pourrait sans doute expliquer la rapidité de l'adaptation (cf Chapitre 2).

#### **Quelles sont l'importance et les éventuelles conséquences des situations de contacts secondaires chez *V. inaequalis* ?**

Les conséquences évolutives d'une transition entre milieu sauvage et agroécosystèmes ne peuvent être correctement abordées qu'en inférant correctement l'histoire démographique des populations. Dans ce manuscrit, nous tenterons de retracer les histoires évolutives de deux types de populations de *V. inaequalis*. Nous développerons l'hypothèse des contacts secondaires dans les deux situations décrites précédemment : 1) l'histoire évolutive des populations virulentes sur pommiers résistants *Vf* (cf Chapitre 1). et, 2) l'histoire évolutive des populations sauvages et d'agroécosystèmes, respectivement CAM et CAP, au Kazakhstan (cf Chapitre 3).

Concernant l'histoire évolutive des populations virulentes sur pommier résistant *Vf*, nous tenterons de proposer plusieurs scénarios alternatifs à l'hypothèse de divergence écologique formulée par Gladieux *et al.* (2011). Notamment, nous testerons l'hypothèse formulée dans Leroy *et al.* (2013) d'une origine sauvage des souches virulentes envers *Vf*.

Concernant les populations kazakhes, une première comparaison de scénarios d'émergence de *Venturia inaequalis* par des méthodes de type Approximate Bayesian Computation (ABC, Tavaré *et al.*, 1997 ; Beaumont *et al.*, 2002) a été réalisée dans l'équipe. Cette étude se base sur l'analyse de plus de 104 marqueurs AFLP et en incorporant l'hypothèse alternative à



l'hypothèse d'host-tracking de contact secondaire entre les populations CAM et CAP. Le scénario évolutif le plus probable est le scénario de contact secondaire ( $P > 0.9$ ). Ainsi, il semblerait que *V. inaequalis* ait émergé probablement par host-tracking lors de la domestication du pommier puis divergé avec la population sauvage puis enfin serait à nouveau entré en contact avec cette dernière population, il y a une centaine de générations (c'est-à-dire environ un siècle). Ce premier résultat semble donc compléter l'hypothèse d'host-tracking précédemment formulée par Gladieux *et al.* (2010) en apportant toutefois une certaine complexité propre aux situations de contacts secondaires.

### **Quelles sont les signatures génomiques de l'adaptation de *V. inaequalis* aux agroécosystèmes**

Le dernier objectif de ma thèse consiste à identifier les zones impliquées dans l'adaptation du pathogène à son hôte domestique et par la même occasion, les modalités évolutives de son émergence dans les agrosystèmes (cf Chapitre 3). L'étude se base sur le re-séquencage du génome de 36 souches. L'histoire évolutive de *V. inaequalis* sera analysée plus en profondeur grâce à l'utilisation de méthodes ABC et basée sur les équations de diffusion  $\delta a \delta i$ .

# Résultats

---

Chapitre 1 :

« Origine de la virulence Vf chez la tavelure du pommier »

p 47

Chapitre 2 :

« Evolution des traits d’histoire de vie de *Venturia inaequalis* en lien avec la domestication du pommier »

p 75

Chapitre 3 :

« Signatures génomiques de l’impact de la domestication du pommier sur son pathogène *Venturia inaequalis* »

p 101



# Chapitre 1

---

## « Origine de la virulence *Vf* chez la tavelure du pommier »

Comprendre les conditions dans lesquelles une nouvelle virulence émerge est crucial afin de gérer et de prévoir l'évolution des maladies qui infectent les plantes. Bien souvent, l'émergence de pathogène est étudié selon un point de vue évolutif, où la nouvelle mutation résulterait du pool de souches non virulentes infectant les hôtes sensibles en réponse au déploiement de nouveaux hôtes résistants.

Dans ce chapitre, nous tentons d'élucider l'origine la plus probable de la virulence *Vf*. Nous avons testé des scénarios utilisant des souches (1) collectées sur *M. floribunda* (2) des populations Européens de souches virulentes et non virulentes (3) ainsi que des souches collectées dans les montagnes d'Asie Centrale (CAM).



## **Take a walk on the wild side : on the origin of the virulent *Vf* strains of the agent of apple scab, *Venturia inaequalis*.**

**Lemaire C\*, De Gracia M.\*, Leroy T., Michalecka M., Lindhard Pedersen H. Gladieux P., Guerin F. and Le Cam B**

\*co first authors

### **Abstract**

Understanding the conditions under which a new virulence emerges is crucial to manage and predict evolution of plant infectious diseases. Theories on emergence of pathogens are often depicted under an adaptive point of view where a new favorable mutation arises from the pool of avirulent strains infecting susceptible hosts in agrosystems in response to deployment of new resistant hosts. In this framework, the role of standing variation from the wild habitat in the potential for adaptation is therefore neglected. In this study we intended to elucidate the most likely origin of a new virulence named *Vf* within the fungus *Venturia inaequalis* that develops gene-for-gene interactions with its host plants *Malus* spp. We tested scenarios using strains collected on the crab-apple *M. floribunda*, the wild progenitor of the corresponding *Vf* resistance gene, European populations of virulent and avirulent strains sampled in orchards on *Malus x domestica* as well as strains collected in Central Asian mountains on *M. sieversii* the ancestor of the domestic apple. Using Approximate Bayesian Computation and bayesian clustering on a microsatellite dataset, we demonstrate that the virulence *Vf* is present in the wild within a population that has diverged from population infecting susceptible hosts since several thousand years. The crab-apple, *M. floribunda* being both the source of genetic resistance used by breeders and the source of virulence, we show that use of wild genetic resources in agrosystems can act as a gateway for virulent population from the wild to invade cultivated habitats. This study demonstrated that non-agricultural environments do provide con-specific strains harboring new virulence lacking in agrosystems. Our findings aims to increase awareness for plant breeders on virulences emergence risks and underline the need to

sample out of agrosystems to evaluate durability of wild genetic resources introgressed in crops.

### **Author summary**

Plant disease resistance relies on a 'gene-for-gene' model where an avirulence effector triggers immunity in plants carrying the matching resistance gene. In this framework, loss of avirulence gene functionality leads to disease. In a control strategy aiming to reduce pesticides in crops using durable resistance varieties, determining origin of virulent strains is a challenge for epidemiologists. Since high evolutionary potential of pathogens, theories on emergence of virulence are depicted under an adaptive point of view where mutations occur in agrosystems and thus neglect the standing variation existing in the wild habitats. Here, we inferred the most likely origin of a new virulence in the fungus *Venturia inaequalis* responsible for the most damaging disease in apple orchards. We showed that virulence originated from the wild and has diverged from strains infecting susceptible varieties in orchards since several thousand years. The crab-apple *Malus floribunda* that is the source of resistance in breeding programs is probably also the source of virulence, the resistant gene would thus act as a gateway for virulence to invade orchards. This study highlights the need to reconsider international policy linked to exploitation of genetic resources in order to limit risks of new virulences emergence associated to human activities.

## INTRODUCTION

Understanding the conditions under which a new virulence emerges is crucial to manage and predict evolution of plant infectious diseases. Causes of such emergence are often depicted under an adaptive point of view. In this framework, it is assumed that hosts harboring new resistance genes induce selective pressures that favor the rise of a new virulence from the pool of avirulent strains. A parallel may be drawn between the recurrent outbreaks of monogenic plant resistances that recognize pathogen effectors (Berngruber *et al.*, 2013) and many infectious diseases *challenging* animal and *human immune* systems (Ma *et al.*, 2009; Nübel *et al.*, 2010). Pathogens are then viewed as quickly evolving organisms, able to rapidly overcome any new resistance in its host. This classical generic model of host-parasite co-evolution, known as the *co-evolutionary arms race* model (Dawkins & Krebs, 1979), dominates interpretations about virulence emergence from field epidemiological to molecular level. Although this model hangs for most of viruses (Forterre & Prangishvili, 2013), the question of its *a priori* generalization is still pending.

Besides, some evidences of contribution to new virulences from wild reservoirs have highlighted the role of standing variation in the adaptive potential of pathogens (Burdon & Thrall, 2008; Leroy *et al.*, 2014; Morris *et al.*, 2009). Adaptations are then supposed to arise by migration, underweighting the apparent rate of adaptive mutation. Unfortunately, examples drawn from plant pathogens in agrosystems are still poorly described (Burdon & Thrall, 2008; Leroy *et al.*, 2014; Morris *et al.*, 2009). Most of epidemiological and evolutionary studies on plant pathogens are restricted to populations sampled on cultivated crops (Morris *et al.*, 2009) due to a lack of knowledge about the actual nature of wild hosts as well as about the characterization of symptoms on non-cultivated hosts.

Agriculture often uses ‘exotic’ plant genotypes (ancient varieties and wild relatives of crop species) as genetic resources for agronomic traits as well as pollinators. For instance resistance genes toward diseases are often introgressed from such a plant material. However, the potential impact of the spatial deployment of these wild plant genetic resources in agrosystems, putatively infected by exotic populations of pathogens remain largely unknown. For example, pathogenicity tests carried out in breeding programs almost always characterize resistance against strains infecting domesticated plants (Lê Van *et al.*, 2012). The potential pathogenicity on agricultural hosts of any population of pathogens adapted to wild relatives is



neglected (Burdon & Thrall, 2014; Zhan *et al.*, 2014). Taking into consideration such wild populations of pathogens in hosts used to introgress resistance in domestic varieties should be of a great interest when one wants resistances to be durable in agrosystems.

Among well-known outbreak of resistance in plants, the emergence of virulent strains of *Venturia inaequalis* against the *Vf* resistance gene in apple is well documented. *V. inaequalis* is the ascomycete responsible for the scab disease on apples (*Malus* spp). Interaction between *V. inaequalis* and *Malus* spp. follows a 'gene-for-gene' model (Flor 1942, 1955, 1971) where a pathogen effector triggers immunity in plants carrying the matching resistance gene (Jones & Dangl, 2006). Seventeen interactions for disease resistance have been described by Bus *et al.* (Bus, Rikkerink, Caffier, Durel, & Plummer, 2011). Among them, the resistance gene *Vf* from the wild relative *Malus floribunda* has been the most introgressed in commercial varieties released by plant breeders. Despite the scarce use of this gene in European orchards, the resistance has been rapidly overcome. The first occurrence of virulence on resistant varieties was reported in Germany in 1988 (Parisi *et al.*, 1993), and thereafter in different other European countries (Bénaouf & Parisi, 2000; Gessler & Pertot, 2012). More, scabbed infected *Malus floribunda* was reported in a private English garden as soon as 1989 even before plantation of *Vf* apple varieties in English orchards (Roberts & Crute, 1994). Using microsatellite variation, first populations genetic studies on French orchards, have shown that virulent strains infecting *Vf* apple varieties were differentiated from strains infecting the non-*Vf* varieties and that they shared a common origin out of sampled orchards (Guérin *et al.*, 2007; Guérin & Le Cam, 2004). More, genetic differentiation between strains sampled on *Vf* and non-*Vf* apple varieties remained strong (Guerin *et al.*, 2007) over several years revealing a low gene flow between them despite presence of the two populations in the same orchards. Gladieux *et al.* (Gladieux *et al.*, 2011) found this reproductive isolation to be congruent with the occurrence of strong exogenous barriers (*i.e* fitness dependent of the environment: Moore and Price, 1993; Kruuk *et al.*, 1999) related to the presence of resistance genes revealed in non-*Vf* apple genotypes. As sexual reproduction in *V. inaequalis* occurs between strains that infect the same leaf, adaptation to host is likely to have strong pleiotropical effects on reproductive isolation (Giraud *et al.* 2006, 2010). In summary, while populations from *Vf* varieties are clearly differentiated and reproductively isolated from populations from non *Vf* ones, an extensive analysis of evolutionary history of both the pathogen and its hosts *Malus* spp is needed to understand the scenario of emergence (new adaptive mutation *vs.* stranding

variation) of virulence *Vf* populations and the maintenance of such a level of genetic differentiation.

Apple Domestication is now well understood (Cornille *et al.* 2014 Trends in Genetics) and evolutionary history of its main pathogen *V. inaequalis* has been shown to be closely related to it (Gladieux *et al.*, 2008). The cultivated apple was domesticated in Central Asia from *Malus sieversii* (Velasco *et al.*, 2010; Cornille *et al.*, 2012). Endemic populations of *M. sieversii* are located in the Tien Shan mountains (border of Kazakhstan, Kirghizstan and Xinjiang Province in China) to the Caspian Sea (Morgan & Richards, 1993) and forms large forests (Harris *et al.*, 2002). It is during the Neolithic period and especially during the development of the Silk Roads that the seeds of wild apple trees were scattered, thus marking the beginning of apple cultures (Juniper & Mabberley, 2006; Wood, 2003, Harris *et al.*, 2002). Domesticated apple would have then spread westward to Europe and eastward to China involved contact and hybridization with closest wild relatives growing along the Silk Route, *M. sylvestris*, *M. baccata*, in Europe, *M. orientalis* in the Caucasus (Janick *et al.*, 1996, Forsline *et al.*, 2003; Cornille *et al.*, 2013; but see also Juniper & Mabberley 2006 for criticisms). *Venturia inaequalis* has followed its host during this history, from mountains of Central Asia to orchards of domestic apples worldwide, through the domestication process (Gladieux *et al.*, 2008; Gladieux *et al.*, 2010; Lê Van *et al.*, 2012). Indeed populations of *V. inaequalis* sampled in *M. sieversii* were identified as the progenies of the ancestral populations of strains infecting domestic apple trees. As for apple trees, pathogen populations coming from the wild *M. sieversii* of Kazakhstan exhibit both genetic and phenotypic differences with the ones sampled in orchards (Gladieux *et al.*, 2010; Lê Van *et al.*, 2012). Indeed, this wild population has diverged about 1000 to 10000 years ago from lineage of *V. inaequalis* found on domestic apple trees, which is congruent with the history of domestication of apple (Gladieux *et al.*, 2010). Concerning *Malus floribunda*, the progenitor of the *Vf* resistance gene it would be a putative hybrid between two wild *Malus* species: *M. baccata* and *M. toringo* respectively endemic of Siberia and Japan (Juniper and Mabberley, 2006) ; *M. baccata* being the carrier of the *Vf* gene (Dunemann *et al.*, 2012; Leroy *et al.* 2013). 10 others *Malus* species endemic in Europe and Asia (Dayton and Williams, 1970). To date, virulence *Vf* has not been detected on *M. sieversii* (Lê Van *et al.*, 2011).

Current knowledge on evolutionary history of both *Malus* spp. and *V. inaequalis* permits to draw four categories of scenarios retracing the putative origin of virulence against *Vf*. In the two first, virulence emergence is explained by a new adaptive mutation within the pool of

avirulent strains, *e.g.* infecting susceptible apple trees in orchards. Two outcomes are then possible. If the same virulent strains would have infected both *Vf* varieties and *M. floribunda*, the progenitor of the *Vf* gene, we then should expect the virulent populations to be poorly differentiated from both avirulent strains sampled in orchard and those sampled on *M. floribunda* (scenario 1). Conversely, if a parallel evolution toward virulence in both *M. floribunda* and *Vf* varieties is assumed, populations from orchards are not expected to be genetically differentiated while they significantly diverge from this sampled on *M. floribunda* (scenario 2). In these two first scenarios, virulence arose from avirulent pool, so both avirulent and virulent strains should derive from the ancestor of Central Asian population on *M. sieversii*. In the third scenario, we hypothesize that *Vf* virulence originates out of the orchards, from *Malus floribunda*. As a result, strains found on *M. floribunda* should (i) appear genetically close to the ones found on *Vf* resistant domestic apple trees and (ii) have diverged from the avirulent population very long time before the introgression of *Vf* resistance into *M. x domestica*. At last in the fourth scenario, virulence against *Vf* could find its origin in the population sampled in Central Asia on *Malus sieversii*. We could hypothesize the virulence gene to segregate in this wild population as we do not know whether *Vf* resistance actually is present in *M. sieversii* populations. We should then expect this population to be divergent from the avirulent domestic one and from the one found on *M. floribunda*.

To elucidate the virulence *Vf* origin, we present here a study using European populations of virulent and avirulent strains sampled in orchards, strains collected in Central Asian mountains on *M. sieversii* as well as strains collected on *M. floribunda*, the wild progenitor of the *Vf* resistance. Using Approximate Bayesian Computation (ABC) (Beaumont, 2002) and bayesian clustering on a microsatellite dataset, we tried to infer the most likely evolutionary scenario of the origin of virulence *Vf* infecting orchards. We will attempt to answer the following question: Do the virulent strains infecting domestic apple in agroecosystems origin from populations found on the wild host *M. floribunda* or do they rather come from an adaptive mutation to virulence within the avirulent population found on susceptible domestic apple trees? Virulence coming from the wild would favor an emergence by migration rather than adaptive mutation in opposition to emergence from avirulent populations, which is the most widely admitted by epidemiologists.

## MATERIALS & METHODS

### Sampling and Genotyping

We studied 821 strains of *V. inaequalis* divided into 14 sites (Table 1). Strains from the *Vf* varieties were collected 5 french localities in 2001, 2003 and 2004 (n=177), in Sweden in 2010 (n = 76), in Poland in 2010 (n = 156) and in Danemark in 2003 (n = 66). Strains from non-*Vf* trees were collected in 4 french localities in 2000, 2003, 2004 (n=186), in Sweden in 2010 (n = 25), in Poland in 2010 (n = 94) and in Danemark in 2003 (n=26). Strains collected on *Malus floribunda* were collected on different European localities from 1989 to 1998. Since they were only 30, we pooled them into a single sample called FLO. We also added two sampling sites collected in Kazakhstan on *Malus sieversii*, the most likely ancestor of the domestic apple tree *Malus x domestica* (Cornille *et al.*, 2012). We pooled the 60 strains into one sample previously called CAM for Central Asia Mountains (see Gladieux *et al.*, 2010).

Genomic DNA was directly extracted from scabbed apple leaf or derived from isolated conidia according to procedure described in Le Cam *et al.*, (2001) and Guérin *et al.* (2007). All samples were genotyped at 8 microsatellite loci (1tc1a, 1tc1b, 1tc1g and 1aac3b from Tenzer *et al.* (1999); Vitc1/2, Vitg11/70, Vicag8/42 from Guérin *et al.* (2004); M42 from Guérin *et al.* (2007) following a method previously described in Guerin *et al.* (2007). As sampling occurred during the asexual phase, identical multilocus genotypes were treated as clones and removed from the analysis. Throughout the study, the terms *virVf*, *avirVf*, CAM and FLO strains are used to make reference to strains sampled on *Vf* apple trees, non-*Vf* apple trees, *M. sieversii* and *M. floribunda*, respectively.

### Diversities and population structure analyses

Classical population genetics parameters like expected *He* and non-biased diversities *Hnb* (Nei, 1978) as well as pairwise *Fst* (Weir & Cockerham, 1984) were estimated using Genetix 4.05 (Belkhir *et al.* 2004). Significant departures from null hypothesis (*Fst* =0) were tested using permutation procedure. Allelic richness was estimated using the method implemented in ADZE (Szpiech *et al.*, 2008). Clustering of strains into populations was performed using the Bayesian model implemented in STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003;

Hubisz *et al.*, 2009). We used a model of admixture with correlated allele frequencies. We set the burn-in length to 500,000 followed by 1,500,000 iterations. For each number of populations *K* set from 2 to 10, five independent runs were performed to check for convergence. We used CLUMPP (greedy algorithm) (Jakobsson & Rosenberg, 2007) to average population membership among the five runs. We used the ad hoc  $\Delta K$  statistics (Evanno *et al.*, 2005) to identify the most relevant number of clusters based on the rate of change in the likelihood of data between successive *K* values. Despite this statistical criterium, other value of *K* may have a relevant biological signification. We will therefore take into account all other clustering that could indicate any other biological information.

**Table 1 : Description of isolates used in the study.**

Localisation	Years	Numbers	VirVf / AvrVf
France	2001 – 2003 - 2004	177	VirVf
Swenden	2010	76	VirVf
Poland	2010	156	VirVf
Danemark	2003	66	VirVf
France	2000 – 2003 - 2004	186	AvrVf
Swenden	2010	25	AvrVf
Poland	2010	94	AvrVf
Danemark	2003	26	AvrVf
<i>Malus floribunda</i> Europe	Between 1989 to 1998	30	
CAM Kazakhstan	2006	60	

## Inference of evolutionary history

In order to find the best scenario of evolutionary history of origin of vir $Vf$  strains we used the Approximate Bayesian Computation method implemented in the software DIYABC (Cornuet *et al.*, 2008). We simulated microsatellite datasets for eight loci corresponding to those used in the observed dataset. Six different models of evolutionary histories were simulated (Fig. 1). We do not use any admixture in models in order to not over parameterize the scenario in face to low number of loci.

We used a Generalized Stepwise Model with two parameters :  $\mu$  the mean mutation rate among loci, and  $P$  the mean parameter of the geometric distribution of the mutation event in terms of number of repeats. Mean mutation rates was drawn from a uniform distribution bounded by extreme value of  $10^{-6}$  and  $10^{-3}$ . Per locus mutation rates were drawn from a gamma distribution (mean =  $\mu$ ; shape = 2). The parameter  $P$  was drawn from a uniform distribution with extreme values 0.1 to 0.3.

In order to compare observed with simulated data, we used the following summary statistics : mean number of alleles per locus, mean diversities, mean size variance, genetics of comparisons between populations ( $F_{st}$ ) and  $d(\mu^2)$  genetic distance.

A polychotomous logistic regression was used to estimate the relative posterior probability of each model using 0.1% of the simulated data sets closest to the observed data. We used the limiting distribution of the maximum likelihood estimators to estimate confidence intervals for the posterior probabilities. We then used a local linear regression to estimate the posterior distributions of parameters under the most likely model. The 0.1% simulated data sets closest to the observed data were used for the regression, after the application of a *logit* transformation to parameter values.

## RESULTS

### Polymorphism and diversities

Analyses of diversities shown in Table 2 indicate that heterozygosities in non-*Vf* samples ( $H_{nb} = 0.67 \pm 0.06$ ) were greater than those of *Vf* samples ( $H_{nb} = 0.38 \pm 0.11$ ), probably reflecting a longer evolutionary history of non-*Vf* populations and/or a larger effective size. Strains sampled on *M. floribunda* showed intermediate heterozygosity of 0.42. Note that CAM sample had a diversity close to non-*Vf* samples ( $H_{nb} = 0.64$ ). Allelic richness is also greater for non-*Vf* samples ( $A = 7.68 \pm 1.89$ ) than for *Vf* ones ( $A = 4.46 \pm 1.92$ ). Samples from *M. floribunda* showed similar allelic richness than *Vf* samples. At last CAM samples exhibited the greatest polymorphism ( $A=11$ , Table 2).

Population structure was also inferred from *Fst* estimates between sampled clustered according to their host of origin (4 groups : vir*Vf*, avir*Vf*, FLO and CAM). All six pairwise *Fst* were significant ( $p<0.001$ ) according to permutation test (1000 permutations). Lowest differentiation was observed between vir*Vf* and FLO group ( $Fst = 0.05$ ,  $p<0.001$ ). Interestingly differentiation estimates were the same for avr*Vf*/vir*Vf* and avr*Vf*/FLO pairwise comparisons ( $Fst = 0.10$ ,  $p<0.001$  for both estimates) as well as for CAM/vir*Vf* and CAM/FLO pairwise comparisons ( $Fst = 0.24$ ,  $p<0.001$  for both estimates).

**Table 2 : Polymorphism and diversity among 8 microsatellite loci at the 14 samples of *Venturia inaequalis*.**

Samples were collected on resistant *Vf* domestic apple trees (named *virVf*, in bold), on non-*Vf* domestic apple trees (named *avrVf*, in italic), on *Malus floribunda* (FLO) and on *M. sieversii* in mountains of Kazakhstan (CAM). He and Hnb are the expected gene haplotype diversity, and non-biased haplotype diversity respectively. A is the allelic richness

Samples	N	He	Hnb	A
<b>PA_VF</b>	22	0.29	0.2975	2.375
<b>LF_VF</b>	79	0.3907	0.3933	5
<b>SW_VF</b>	76	0.22	0.2215	3.25
<b>PL_VF</b>	156	0.549	0.5509	8
<b>DK_VF</b>	66	0.4241	0.4276	3.6667
<i>PA_avrVF</i>	73	0.7176	0.7227	11.25
<i>LS_avrVF</i>	30	0.7031	0.7157	7
<i>LF_avrVF</i>	57	0.7153	0.7218	9
<i>YV_avrVF</i>	26	0.6703	0.6839	6.25
<i>PL_avrVF</i>	94	0.709	0.7138	10.375
<i>DK_avrVF</i>	26	0.5055	0.5158	4.6667
<i>SW_avrVF</i>	25	0.6319	0.6455	5.25
FLO	30	0.4129	0.4204	3.625
CAM	60	0.6405	0.6462	11

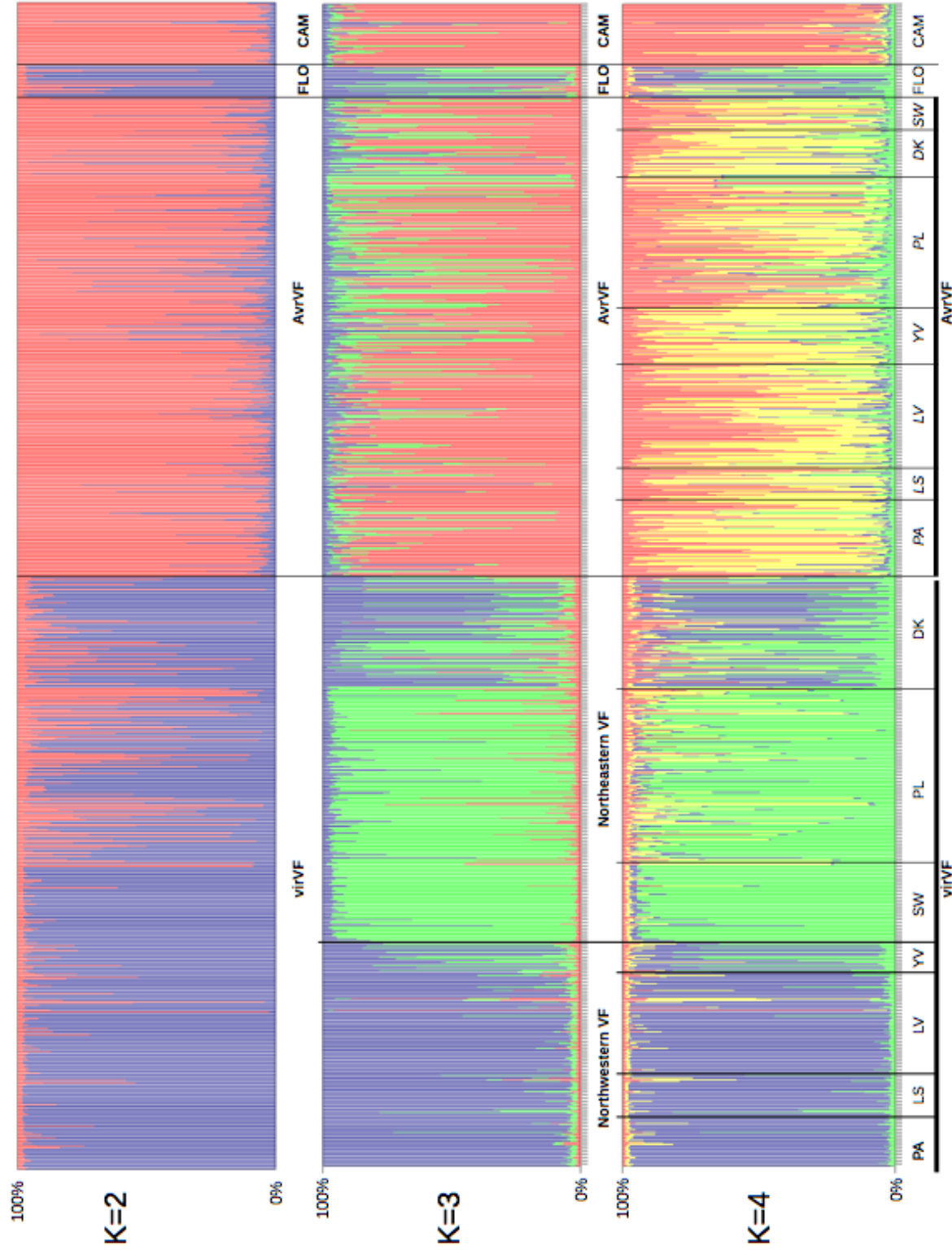


## Population clustering

Bayesian clustering approach using STRUCTURE revealed a partition between strains coming from *M. floribunda* and *Vf Malus* on one side and strains sampled on non-*Vf* and wild Central Asia Mountains (CAM) *Malus* on the other side (Fig. 2). The clustering of strains into two groups ( $K=2$ ) was the most likely according to Evanno's test (Evanno *et al.*, 2005). Although less likely, higher  $K$  may reveal relevant genetic structure and may therefore give interesting biological insights. Thus using  $K = 3$ , the vir $Vf$  strains were divided into two groups according to their geographical origins : samples coming from Northeastern Europe (Denemark, Poland, Sweden) differentiated from those sampled in Northwestern Europe (Fig. 2b). The FLO strains clustered with the Northwestern Europe group of vir $Vf$  strains. STRUCTURE then differentiated CAM samples from avr $Vf$  strains coming from European domestic apple trees only when using  $K=4$ . This result suggests a more recent common ancestry between these strains.

## Evolutionary scenario

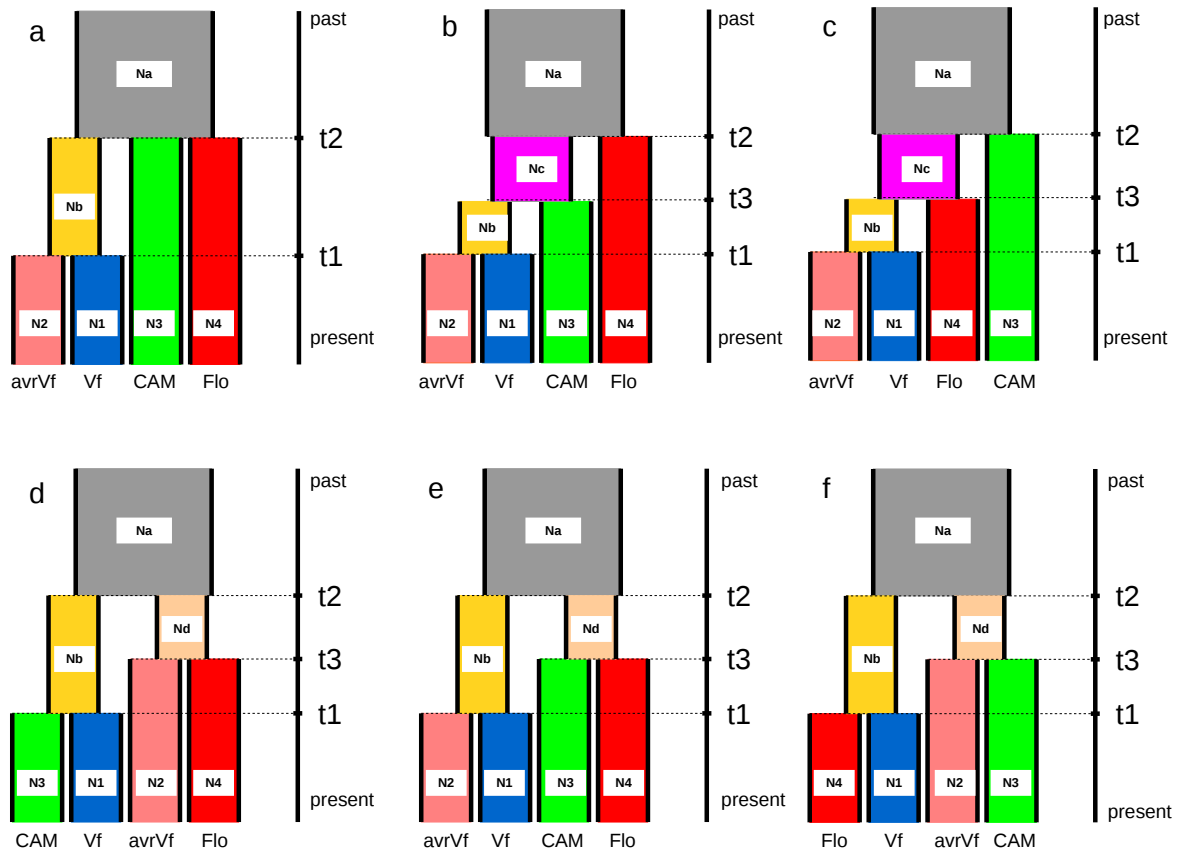
Among the six evolutionary scenarios implemented in DIYABC (Cornuet *et al.*, 2008), the model 6 is the best supported one. Indeed, using a logistic regression approach to estimate the relative posterior probabilities of each scenario, strongest statistical support was provided for the model 6 (posterior probability,  $p = 0.51$ , 95% confidence interval 0.5031-0.5180) while other models did not provide posterior probabilities greater than 0.2. This scenario assumes an ancient divergence between strains sampled on *Vf*-trees and *M. floribunda* on one side, and strains sampled on non-*Vf* and *M. sieversii* trees on the other side. Model 2 assuming that strains from *M. floribunda* diverged first followed by the divergence of the strains from *M. sieversii* in Kazakhstan mountains provided lowest support (posterior probability,  $p = 0.02$ , 95% confidence interval 0.0185,0.0221). Changing the order of divergence between CAM and *M. floribunda* strains (model 3) provided a greater support (posterior probability,  $p = 0.17$ , 95% confidence interval 0.1672,0.1782) but remains far lower than scenario 6. Then all parameters were estimated from the scenario 6. Posterior densities of parameter estimates of divergence times as well as effective sizes are large (Table 3). The nature and the low number of markers did not permit us to get more precise parameters estimates. Although rough estimate, 95% posterior distribution divergence time  $tI$  between vir $Vf$  and FLO strains ranged from *ca.* 70 to



**Figure 2: Proportion of ancestry of *Venturia inaequalis* haplotypes from  $K=2$  to  $K=4$  clusters inferred with the Structure software.**

Each vertical bar representing an individual is partitioned in  $K$  segments representing the proportion of its assignment to the  $K$  clusters.

between CAM and *avrVf* strains (*i.e.* European avirulent strains on *M. x domestica*) ranged from *ca.* 1900 to 4000 years BP, that matches posterior estimates proposed by Gladieux *et al.* (2010) under an isolation with migration (IM) model. The most ancient divergence occurred between the clade containing *avrVf* and FLO strains on one side, and the one containing *avrVf* and CAM strains on the other side and was estimated to *ca.* 24000 years. Effective sizes of each population has a mode of the order of  $10^4$ .



**Figure 1 : Evolutionary scenarios compared by Approximate Bayesian Computation:** Model assumes that both *virVf* and *avrVf* directly derived from a common ancestor (Nb), CAM and FLO populations have independent lineages. Model b assumes that CAM and the ancestor of *VirVf* and *avrVf* populations (Nb) have the same ancestor (Nc) and FLO is a separate lineage. The model c is similar to model b but CAM and FLO lineages were permuted. In scenario d, CAM and *virVf* population derived from Nb, *avrVf* and FLO derived from Nd, and both Nb and Nd derived from Na. Model e assumes that *avrVf* and *virVf* derived from the ancestor Nb, and CAM and Flo derived from the ancestor Nd. At last model f assumes that Flo and *virVf* share the same ancestor Nb and *avrVf* and CAM share Nd as ancestor. Nx denote the effective sizes of population x. Divergence times are noted tx.

**Table 3 : Posterior distributions of demographic and mutation parameters estimated by Approximate Bayesian Computation for model 6.**

For each parameter mode and 95% boundaries (CI2.5-CI97.5) of the posterior distribution are given.  $N_{virVf}$ ,  $N_{avrVf}$ ,  $N_{CAM}$  and  $N_{FLO}$  are effective sizes of *virVf*, *avrVf*, CAM and FLO respectively. Parameters  $t1$ ,  $t2$ , and  $t3$  are divergence times between *virVf* and FLO, CAM and *avrVf* and (*virVf*-FLO) and (*avrVf*-CAM) respectively

Parameter	mode	CI2.5	CI97.5
$N_{Vf}$	3.63e+04	9.19e+03	6.32e+04
$N_{non-Vf}$	6.85e+04	2.49e+04	1.21e+05
$N_{CAM}$	8.60e+04	2.72e+04	1.70e+05
$N_{FLO}$	2.53e+04	1.10e+04	6.74e+04
$t1$ <i>Vf</i> -FLO	6.70e+01	1.57e+01	2.47e+02
$N_a$	2.52e+04	4.39e+03	1.90e+05
$t2$ non- <i>Vf</i> -CAM	8.99e+03	1.94e+03	3.99e+04
$N_b$	7.56e+05	2.02e+05	1.46e+06
$t3$ ( <i>Vf</i> -FLO)-(non- <i>Vf</i> -CAM)	2.37e+04	8.58e+03	5.01e+04
$N_c$	2.45e+04	3.09e+03	1.45e+05
$\mu_{mic\_1}$	4.84e-05	2.10e-05	1.65e-04
$p_{mic\_1}$	4.00e-01	2.12e-01	5.80e-01

## DISCUSSION

Emergences of new virulences are often depicted as a full adaptive process where a new favorable mutation arises from the pool of avirulent strains in agrosystems in response to the occurrence of new resistant hosts. The role of standing variation in the potential for adaptation was therefore neglected. However most of pathogens have a large distribution and may colonize a larger range of hosts than cultivated ones. One interesting property of adaptation from standing variation is that it involves migration/recombination which is more likely than adaptive mutation (Karasov *et al.*, 2010). Migration from non-agricultural hosts toward a domestic one, is particularly likely as wild relatives are often used for introgression of new resistances and pollination of auto-incompatible species. Our study showed that the virulence against the *Vf* resistance did not arise from the pool of avirulent strains infecting non-*Vf* apple trees in orchards, but rather originated from the wild. The crab apple *Malus floribunda* that is the source of *Vf* resistance in breeding programs, is probably also the source of the virulence. Our results have three major consequences. First the use of wild genetic resources of resistance in agrosystems can act as a gateway for virulent population from the wild. How should breeders then impede migration of preexistent virulent pathogens on wild hosts toward new domestic resistant ones? Second consequence addresses the question of secondary contacts in pathogens which diverged since several ten thousands of years. As such event can boost genetic diversity of crop pathogens, it may promote the evolution of more aggressive hybrids and therefore may have epidemiological consequences (Leroy *et al.* 2014). Third, this finding makes us to address the question of the nature of genomic signatures of new virulences proceeding from standing variation rather than new favorable mutation.

### The *Vf* virulent strains emerged from a crab-apple tree

The ornamental crab-apple tree, *Malus floribunda* carrying the *Vf* resistance gene was first used in 1914 as a source parent for the complete resistance to control scab in orchards. Approximate Bayesian Computation and Bayesian clustering permitted to find that vir $Vf$  strains more likely occurred from *M. floribunda* than from common non-*Vf* varieties. Indeed, virulent populations sampled on *Vf* varieties showed almost the same values of genetic parameters than strains coming from *M. floribunda*. In addition, the retained model permitted to exclude any possibility for any divergence of virulent strains from avirulent populations followed by secondary migration of virulent strains toward *M. floridunda*. In fact divergence

and differentiation between *virVf*/FLO and *avrVf*/CAM lineages is too high (divergence time  $t = 24000$  years, 95% CI 8600-50,000) for this event to be compatible with the contemporary history of introduction of the resistant *Vf* varieties in orchards. Indeed, first observations of virulent *Vf* strains occurred in 1988 in a collection of scab resistant *Malus* species in Germany (Parisi *et al.*, 1993) and in 1989 in the United Kingdom in a private garden on *Malus floribunda* trees (Roberts and Crute, 1994). More, identification of the two mating types infecting within this tree even prior the use of *Vf* varieties in English orchards strongly suggest preexistence of this cryptic population out of agrosystems. In our ABC approach, we did not test models with admixture between groups since the low number of loci would not permit us to estimate clearly overparameterized models. Nevertheless, no migration or admixture was detected between *virVf* and *avrVf* strains in several french orchards that were monitored over years (Gladieux *et al.*, 2011).

Epidemiological studies mostly described diseases on the cultivated hosts and in the rare cases where wild relative hosts are studied, they are often infected by a relative species of pathogen (Morris *et al.*, 2009). Our study demonstrates that non-agricultural environments do provide conspecific strains harboring different virulences, and so showed the necessity to sample out of agrosystems to elucidate the origin of new virulence emergences.

#### Use of wild genetic resources as a cause of new virulences emergence in agrosystem?

Emergence of diseases related to introduction of exotic pathogens in new area due to international trade is well known and documented (Fisher *et al.*, 2012). In this study, we describe a new process of emergence *ie* introduction in cultivated habitat of a new virulence from the wild facilitated by the introgression of a resistance gene. The introgression into varieties of the resistance gene *Vf* carried by the crab-apple *M. floribunda* would have then acted as a gateway for virulent strains present on wild species to infect orchards. In summary *Vf* resistance gene in apple trees could be so considered as a compatibility gene for a wild population to invade agrosystem. This finding puts the question of the potential risk for wild genetic resources to provide new virulences, and consequently to strongly weaken the durability of newly introgressed resistance traits. Indeed, once introduced in agrosystem, theoretical studies suggest that this generalist population may become invasive (Sapoukhina *et al.*, 2009) provided that mating with endogenous population is possible. More generally, extensive studies of pathogens infecting wild hosts should permit to identify eventual other

populations of conspecific pathogens characterized by new virulence that would threaten crops. As the key factor to control new emergence is limiting the occurrence of gene flow between strains infecting a wild genetic resource and a new host carrying resistance introgressed from this wild relative, it should be relevant to create conditions favouring isolation of these new resistant hosts (Leroy *et al.*, 2014). For instance, prohibiting the implantation of resistant hosts close to geographic areas of wild host should be a prerequisite. Overall in a context of an intensive international trading and an increasing deforestation, this study raises the question of gene flow consequences between the wild and cultivated habitats in the emergence of new pathogens.

### Evolutionary consequences of secondary contacts?

Wild origin of virulent strains of *V. inaequalis* has a major consequence in terms of potential for reproductive isolation with agrosystems populations. Previously, Gladieux *et al.* (2011) did not find any evidence of gene flow between sympatric populations of virulent and avirulent strains of *V. inaequalis*. This was depicted as an evidence of strong effect of adaptation to host causing pleiotropically reproductive isolation reinforced by the occurrence of exogenous barriers also existing in non*Vf* varieties. Here we show that virulent strains come from a wild host and have probably diverged in allopatry from lineage of avirulent strains for several thousands of years (*ca.* 24000 years). Our finding renews the discussion by giving more emphasis to migration at expense of adaptive mutation. More, migration of virulent strains from their wild host to orchards on resistant varieties mixed with susceptible ones is comparable to a secondary contact between divergent populations. According to these results could we conclude to a complete reproductive isolation? If so, are barriers to reproductive isolation exclusively ecological? In the case of a wild origin of new virulences, reproductive isolation may arise from the action of several barriers linked or not to ecology (Barton & Hewitt, 1985; Kruuk *et al.*, 1999; Bierne *et al.*, 2011). Answers to this question depend on the length of allopatric divergence before secondary contact. In fact, during this period, mutations accumulate in each lineage. When secondary contact occurs, some mutations at different loci in different lineages may negatively interact in hybrids and give intrinsic incompatibilities described by Dobzhansky and Muller (DMI). Incompatibilities are not necessarily related to ecological factor but are involved in hybrid non-viability or sterility. Given that *virVf* and *avrVf* populations are now in secondary contact in apple orchards, such

incompatibilities may be associated with ecological factors and contribute together to the reproductive isolation observed in Gladieux *et al.* (2011).

At last, secondary contacts may lead to partial rehomogenization and hybrid zones (Barton & Hewitt, 1985). The fate of a secondary contact is driven by the equilibrium between gene flow and selection against alleles coming from the other population. At this time we cannot predict how this secondary contact will evolve. During the coming years, monitoring and researching hybridization events consecutive to these secondary contacts will be investigated in these different orchards. Evaluating pathogenicity traits of life of the hybrids progenies will be critical to gain further insights into the evolution of this unprecedented process of virulence emergence.

#### Finding virulences in genomes : hard or soft sweeps?

New virulent emerging strains are supposed to experience strong demographic expansions as they can invade empty habitats. Classically mutation responsible of virulence is assumed to fix in populations infecting resistant hosts. In the “mutation-toward-virulence” model described in Leroy *et al.* (2014), virulence appears after the use of resistant hosts. We then expect a strong decrease of genetic diversity in the flanking region involved in the virulence, a process called hard sweep (Maynard-Smith & Haigh, 1974; Hermison & Pennings, 2005; Penning & Hermison, 2006). This zone of strong differentiation between virulent and avirulent strains is then called genomic island of differentiation (Turner *et al.*, 2005) or genetic barrier of reproductive isolation (Barton & Gale, 1993; Bierne *et al.*, 2011). In our model of “virulence from standing variation”, the mutation responsible of virulence preexists to the apparition of resistant hosts. In this case, favorable mutation segregates in the population and freely recombines with surrounding polymorphisms. At the time of selection frequency of the favorable mutation will increase. However, polymorphism in the flanking regions will not decrease as much as for *de novo* mutations, due to previous recombination. This process is known as soft sweep (Hermison & Pennings, 2005; Penning & Hermison, 2006, Messer & Petrov 2013). According to Messer and Petrov (2013) soft sweeps are more likely to occur than hard sweeps and should be pervasive in genomes. Thus genomic region of *V. inaequalis* involved in the *Vf* virulence is supposed to be associated with a soft sweep rather than a hard sweep. Indeed, flanking polymorphism may have segregated around the *Vf* mutation. Figure 2 shows the individual probabilities of assignment to a cluster using  $K=3$ .



Virulent strains are assigned to different populations according to their geographical localities. Indeed Northeastern European populations are different from Southwestern european ones. This result could suggest different origins of virulent strains from other *M. floribunda* or even other putative wild *Malus* carrying the *Vf* gene. To date the corresponding avirulence locus *AvrVf* is not yet cloned in *V. inaequalis*. Once identified it should permit to access to allelic variation at this locus within populations of *V. inaequalis* in cultivated and wild habitat and to test hypothesis of a soft sweep around the gene of interest.

## CONCLUSION

Genetic resources of valuable traits for plant breeders become rare and their introgression in varieties is time consuming especially for perennial crops. Leroy et al (2014) underlined necessity to protect new introgressed resistance from virulences emerging from the wild by taking severe sanitary measures in worldwide trades, importation of wild resources from their center of diversity and development of optimal strategies for spatial resistance deployment. During our European surveys we did not observe over years any new virulence emergence from avirulent populations infecting orchards In the same line, *Vf* breakdown in commercial orchard has never been reported in countries out of Europe (Gessler & Pertot, 2012) ; a scab infection was just noted in 2003 and 2005 on a *M. floribunda* tree in an arboretum in USA (Ohio state) but this tree has never been unequivocally identified as being the accession 821 (Beckerman *et al.*, 2009). This strongly suggests that the gene involved in *Vf* avirulence trait may have an essential role in the life cycle of the pathogen and that its alteration would strongly affect its fitness preventing it to emerge within the pool of avirulent strains. We might then hypothesize that life expectancy of the *Vf* resistance gene will be much higher than supposed especially in apple cultivated areas with no wild endemic *Malus* trees harboring the *Vf* resistance that could host virulence. Indeed as all breeding programs view the selection for scab disease resistance as a priority, optimal use of this resistance gene should greatly help to reduce fungicide use in very important apple producers countries in Oceania, South Africa and South America.

Overall in a context of an intensive international trading and an increasing deforestation, this study raises the question of gene flow consequences between the wild and cultivated habitats in the emergence of new pathogens.

## ACKNOWLEDGMENTS

The authors wish to thank Dr. Nylbom for providing samples from Sweden. This work was supported by INRA—SPE, Region Pays de La Loire, Ministère de l'Éducation Nationale de la Recherche et de la Technologie

## REFERENCES

- Barton NH & Hewitt GM. 1985** Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113-148.
- Beaumont MA, Zhang W, Balding DJ. 2002.** Approximate Bayesian Computation in population genetics. *Genetics* **162**: 2025-2035.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004.** Genetix 4.05. Université de Montpellier II, France.
- Benaouf G, Parisi L. 2000.** Genetics of host-pathogen relationships between *Venturia inaequalis* races 6 and 7 and *Malus* species. *Phytopathology* **90**: 236-242.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011.** The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044-2072.
- Beckerman J, Chatfield J, Draper E. 2009.** A 33-year Evaluation of Resistance and Pathogenicity in the Apple Scab—crabapples Pathosystem. *HortScience* **44**, 599–608.
- Bus VG, Rikkerink EH, Caffier V, Durel CE & Plummer KM. 2011.** Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*. *Annual Review of Phytopathology* **49**: 391-413.
- Burdon JJ, Thrall PH., 2008.** Pathogen evolution across the agro-ecological interface: implications for disease management. *Evolutionary Applications* **1**: 57–65.
- Cornille A, Gladioux P, Smulders MJ, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L *et al.* 2012.** New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* **8**: e1002703.
- Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ,**

- Guillemaud T, Estoup A 2008.** Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computation. *Bioinformatics* **24**: 2713-2719.
- Cornille A, Giraud T, Bellard C, Tellier A, Le Cam B, Smulders MJ, Kleinchmit J, Roldan-Ruiz I, Gladieux P. 2013.** Postglacial recolonization history of the European crab apple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* **22**: 2249-2263.
- Cornille A, Giraud T, Smulders MJ, Roldan-Ruiz I, Gladieux P. 2014.** The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**: 57-65.
- Dawkins R & Krebs JR. 1979.** Arms races between and within species. Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain), **205**: 489–511
- Dayton DF & Williams EB. 1970.** Additionnal allelic genes in *Malus* sp. for scab resistance of two reaction types. *J Am Soc Hortic Sci.* **92**: 89-93
- Dunemann F, Gläss R, Bartsch S, Saad Eldin M, Peil A, Bus V. 2012.** Molecular cloning and analysis of apple *HcrVf* resistance gene paralogs in a collection of related species. *Tree Genetics & Genomes* **8**: 1095-1109.
- Evanno G, Regnaut S, Goudet J. 2005.** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611-2620.
- Falush D, Stephens M, Pritchard JK. 2003.** Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.
- Fisher MC, Henk D, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012.** Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484** : 186–94.
- Flor HH. 1942.** Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology* **32**: 653–669.
- Flor HH. 1955.** Host-parasite interaction in flax rust - its genetics and other implications. *Phytopathology* **45**: 680–685.
- Flor HH. 1971.** Current status of the gene-for-gene concept". *Annual Review of Phytopathology* **9**: 275–296.
- Forsline PL, Aldwinckle HS, Dickson EE, Luby JL, Hokanson SC. 2003.** Collection,

maintenance, characterization and utilization of wild apples of central Asia. *Horticultural Reviews*, 29 DOI: 10.1002/9780470650868.ch1

**Forterre P & Prangishvili D. 2013.** The major role of viruses in cellular evolution: Facts and hypotheses. *Current Opinion in Virology* **3**: 558–565.

**Gessler C & Pertot I. 2012.** *Vf* scab resistance of *Malus*. *Trees-structure and function* **26**: 95-108.

**Giraud T, Villaréal LMMA, Austerlitz F, Le Gac M, Lavigne C. 2006.** Importance of the lifecycle in sympatric host race formation and speciation of pathogens. *Phytopathology* **96**: 280-287.

**Giraud T, Gladieux P, Gavrillets S. 2010.** Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution* **25**: 387-395.

**Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, Didelot F, Le Cam B. 2011.** Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology* **20**: 4521-4532.

**Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B. 2008.** On the origin and spread of the scab disease of apple: out of Central Asia. *PloS One* **1**: e1455.

**Gladieux P, Zhang XG, Roldan-Ruiz I, Caffier V, Leroy T, Devaux M, Van Glabeke S, Coart E, Le Cam B. 2010.** Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology* **19**: 658-674.

**Guerin F and Le Cam B. 2004.** Breakdown of the scab resistance gene *Vf* in apple leads to a founder effect in populations of the fungal pathogen *Venturia inaequalis*. *Phytopathology* **94**: 364-369.

**Guerin F, Gladieux P, Le Cam B 2007.** Origin and colonization history of newly virulent strains of the phytopathogenic fungus *Venturia inaequalis*. *Fungal Genetics and Biology* **44**: 284-292.

**Harris SA, Robinson JP, Juniper BE. 2002.** Genetic clues to the origin of the apple. *Trends in Genetics* **18**: 426-430.

**Hermisson J & Pennings PS. 2005.** Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169** : 2335–2352.

- Pennings PS & Hermisson, J. 2006.** Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution* **23**: 1076–1084.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009.** Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**: 1322-1332.
- Janick J, Cummins JN, Brown SK & Hemmat M. 1996.** Apples. In: *Fruit Breeding* (eds. Janick J & Moore JN). Wiley, New York.
- Jakobsson M & Rosenberg NA. 2007** CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**:1801-1806.
- Juniper BE & Mabberley DJ. 2006.** *The story of the apple*. Timber Press, Portland.
- Jones JD & Dangl JL. 2006.** The plant immune system. *Nature* **444**: 323-329.
- Karasov T, Messer PW, Petrov D A. 2010.** Evidence that Adaptation in *Drosophila* Is Not Limited by Mutation at Single Sites. *PLoS Genetics* **6**: e1000924
- Kruuk LE, Baird SJ, Gale KS, Barton NH. 1999.** A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**: 1959–1971.
- Le Cam B, Devaux M, Parisi L. 2001.** Specific polymerase chain reaction identification of *Venturia nashicola* using internally transcribed spacer region in the ribosomal DNA. *Phytopathology* **91**: 900-904.
- Leroy T, Le Cam B, Lemaire C. 2014.** When virulence originates from non-agricultural hosts: new insights into plant breeding. *Infection, Genetics and Evolution*. doi:10.1016/j.meegid.2013.12.022.
- Leroy T, Lemaire C, Dunemann F, Le Cam B. 2013.** The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology*. **13**:doi:10.1186/1471-2148-13-64.
- Lê Van A, Durel CE, Le Cam B & Caffier V. 2011.** The threat of wild habitat to scab resistant apple cultivars. *Plant Pathology* **60**: 621-630

- Lê Van A, Gladieux P, Lemaire C, Cornille A, Giraud T, Durel CE, Caffier V, Le Cam B. 2012.** Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications* **5**: 694-704.
- Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B. 2010.** Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367-373.
- Maynard Smith J & Haigh J. 1974.** The hitch-hiking effect of a favourable gene. *Genetical Research* **23**: 23-35.
- Morgan J & Richards A. 1993.** *The book of apples*. The Ebury Press, London.
- Moore WS & Price JT. 1993.** Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. Pp. 196-255 in R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, New York.
- Morris CE, Bardin M, Kinkel LL, Moury B, Nicot PC, Sands DC. 2009.** Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. *PloS Pathogens* **5**: e1000693.
- Nei M 1987.** *Molecular evolutionary genetics*. Columbia University Pr.
- Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla S *et al.*, 2010.** A Timescale for Evolution, Population Expansion, and Spatial Spread of an Emerging Clone of Methicillin-Resistant *Staphylococcus aureus*. *PLoS Pathogens* **6**: e1000855.
- Parisi L, Laurens F, Guillaumes J, Kruger J. 1993.** A new race of *Venturia inaequalis* virulent to apples with resistance due to the *Vf* gene. *Phytopathology* **83**: 533-537
- Pritchard JK, Stephens M, Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.
- Roberts AL & Crute IR. 1994.** A scab resistance from *Malus floribunda* 821 (*Vf*) is rendered ineffective by isolates of *Venturia inaequalis* from *Malus floribunda*. Norwegian Journal of Agricultural Science **17**: 403-406
- Sapoukhina N, Durel CE, Le Cam B. 2009.** Spatial deployment of gene-for-gene resistance governs evolution and spread of pathogen populations. *Theoretical Ecology* **2**: 229–238.

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, 24, 2498–2504. doi:10.1093/bioinformatics/btn478

**Tenzer I, Degli Ivanisovich S, Morgante M, Gessler C. 1999.** Identification of microsatellite markers and their application to population genetics of *Venturia inaequalis*. *Phytopathology* **89**: 748-753.

**Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. 2010.** The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**: 833-939.

**Weir BS & Cockerham CC. 1984.** Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

**Wood F. 2003.** *The Silk Road*. The Folio Society, London.

## Chapitre 2

---

### « Evolution des traits d'histoire de vie de *Venturia inaequalis* en lien avec la domestication du pommier »

Dans le chapitre précédent, nous montrons que le pommier *M. floribunda* est à la fois la source de la résistance génétique utilisée par les éleveurs et la source de virulence. Cette étude a montré une conséquence éventuelle des situations de contact secondaire et l'impact que peut avoir le développement de l'agriculture et les pratiques culturelles.

La disponibilité des ressources génétiques et des connaissances précieuses sur le pathosystème fournit un excellent cadre pour étudier l'impact de la domestication de l'hôte sur l'évolution des traits d'histoire de vie des agents pathogènes. Il a récemment été démontré que la coévolution, ou host-tracking, entre le pommier et la tavelure a conduit à des changements dans la virulence et à une augmentation de l'agressivité des souches (Lê Van *et al.*, 2012).

Dans ce chapitre, d'autres caractéristiques phénotypiques ont été analysées afin de démontrer la divergence des deux populations CAM et CAP. Les traits qui ont été choisis sont impliqués dans la dispersion et la colonisation : la croissance mycélienne, le taux de germination, la taille et le nombre des spores asexuées. Nous cherchons ainsi à définir les implications d'un éventuel changement phénotypique sur l'histoire des traits d'histoire de vie de *V. inaequalis* dans les agrosystèmes.





## How did host domestication modify life history traits of its pathogens?

De Gracia M.<sup>1,2,3</sup>, Cascales M.<sup>1,2,3</sup>, Expert P.<sup>1,2,3</sup>, Bellanger MN.<sup>1,2,3</sup>, Le Cam B\*.<sup>1,2,3</sup>, Lemaire C\*.<sup>1,2,3</sup>

<sup>1</sup> INRA, IRHS, SFR QUASAV, Rue Georges Morel, Beaucouzé, 49071 France

<sup>2</sup> Université d'Angers, IRHS, PRES UNAM, SFR QUASAV, Boulevard Lavoisier, Angers, 49045 France

<sup>3</sup> Agrocampus Ouest, IRHS, SFR QUASAV, Rue Le Nôtre, Angers, 49045 France

\*\*co-senior authors

### Abstract

Understanding evolutionary dynamics of the pathogens during the domestication of their hosts and the rise of agroecosystems is essential to bring disease-causing agents under durable human control. Here, we investigate the changes in life-history traits of the fungal pathogen *Venturia inaequalis* during the domestication of apple. Life traits linked to the fungus dispersion were compared between 60 strains sampled in domestic and wild habitats in Kazakhstan, the center of origin of both host and pathogen. Our two main findings are that the transition from wild to agroecosystems was associated with an increase of the average of both spore sizes and sporulation capacity and that distribution of each quantitative trait of the domestic population mostly overlapped with those of the wild population. Our results suggest that apple domestication had considerable impact on fungal characters linked to its potential for dispersal by selection from standing phenotypic diversity. We show that pestification of *V. inaequalis* in orchards led to an enhanced allocation in colonization ability for which potential capacity did already exist in the wild area. This study emphasizes the potential threat of pathogens populations living in wild environment for durability of resistance in crops.

**Keywords :** life-history traits, domestication, fungus, pathogen, *Venturia inaequalis*, dispersal

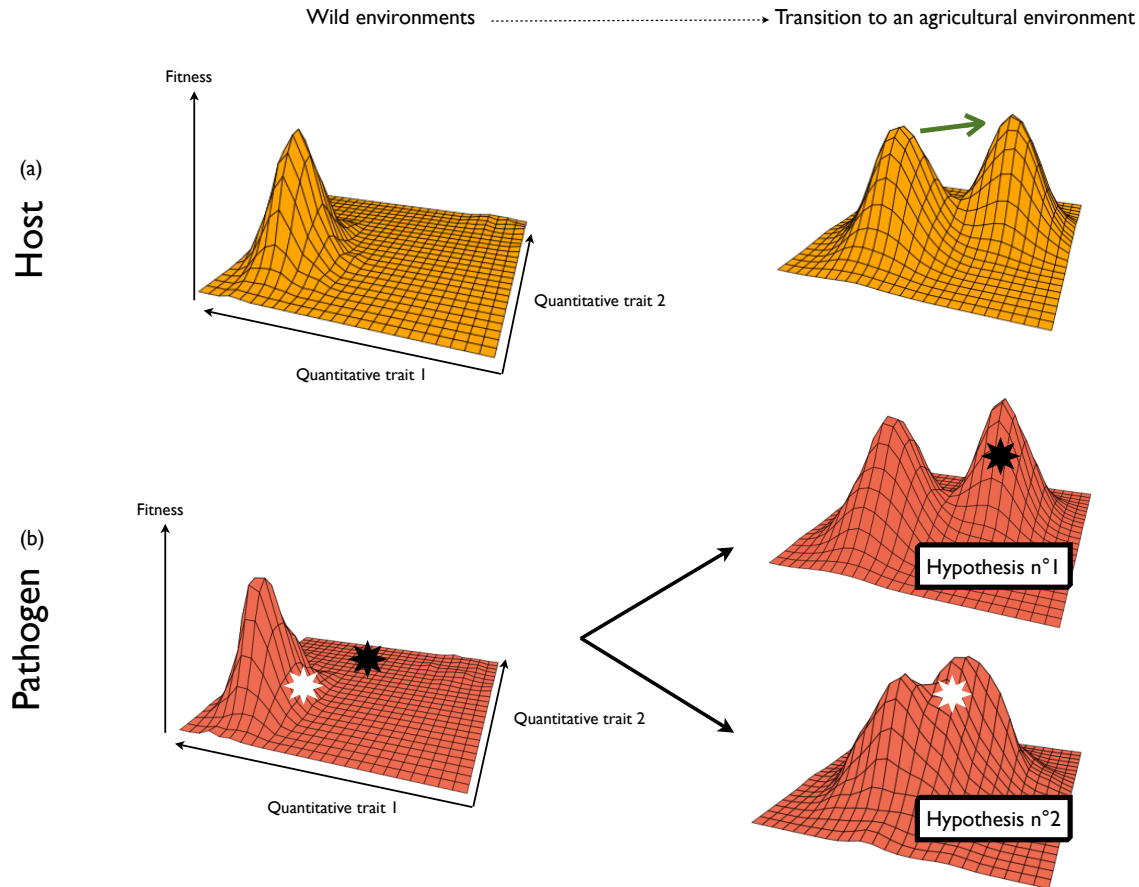
## INTRODUCTION

While the impact of domestication is now well described on many organisms (Tenaillon *et al.*, 2004; Liu *et al.*, 2006; Haudry *et al.*, 2007; Glémin & Bataillon, 2009) little attention has been paid to the evolutionary dynamics of the pathogens during the domestication of their respective hosts and the rise of agroecosystems (Buckler, 2001; Stukenbrock & McDonald, 2008). Yet understanding how life history traits of pathogens were affected through the crucial process of domestication of their host is particularly relevant to figure out which traits were selected for pathogens to adapt to their new environment. The domestication of plants and animals is one of the most important advances of the past 13,000 years in agriculture. It can be defined as an evolutionary process triggered by the anthropogenic selection of individuals from wild populations based on agronomic criteria (Clement, 1999). Continuous selection of wild species to new anthropic environments like agroecosystems has governed our current diet, the rise of civilizations and the world population demography; mixing a large community of disciplines interested in the subject. So anthropic selection based on some phenotypic traits resulted in a strong accumulation of fixation of alleles, some of them being adaptive, in the genomes of domesticated plants (Buckler, 2001; Hyten *et al.*, 2006; Glémin & Bataillon, 2009) animals (Rubin *et al.*, 2010) but also microorganisms (Liti *et al.*, 2009). The set of morphological and/or genetic differences between domesticated species and their wild progenitors defines the domestication syndrome (Hammer, 1984; Li *et al.*, 2006). Among pathogens attacking our basic food resources, fungi cause more than a third of emerging diseases (Anderson *et al.*, 2004; Giraud *et al.*, 2010). Deciphering the way by which pathogens living on wild hosts adapt to agricultural constraints would allow us to manage emergences of diseases and improve the sustainability of genetic resistances of hosts.

As domestication of crops is viewed as a long and continuous process that probably spans over thousands years, the subsequent phenotypic evolution of the associated pathogen is then expected to be continuous as well. This model of parallel divergence of host and its pathogen is known as host-tracking (Stukenbrock & McDonald, 2008) and assumes a close dependence of the pathogen to the phenotypic value of its host. Indeed, pathogens can emerge on domesticated hosts by following its host along the process of domestication (Gladieux *et al.*, 2010; Gifford, 2012). In this way, a disruptive selection occurs between wild and cultivated forms, both in the host and in the pathogen. A dichotomy has been observed on certain quantitative traits: certain traits are favored in the wild, some in the domestic environment (Abbo, 2014). In fact, the efficiency of selection toward evolution of pathogen

depends on its phenotypic diversity. The question here is how much does host domestication change the adaptive landscape for the pathogen. Although domestication creates new phenotypes of hosts, it does not necessarily change what matters for the pathogen-resistance genes, for instance. During the domestication of plants, some genotypes of crops have been selected on the basis of phenotypic characters of agronomic interest. For example, when the first sedentary human populations arose, the seed harvest has been conditioned by certain traits of agronomic interest, like seed size, germination rate, quality seed or even traits related to flowering time (Buckler, 2001). Selection for resistance to diseases may not have been the target of the first farmers. Phenotypic changes of pathogens, if any, should have been triggered "unconsciously".

Beside the proper phenotypic evolution of the plant during domestication, the influence of transition "natural environment / domestic environment" in the case of host-pathogen co-evolution is not negligible. Agroecosystems designed by human, are special structures with characteristics favoring emergence of pathogens. The development of agriculture favored the ability for plants to grow in new habitats with higher densities, but less genetic diversity, increased and uniform nutrient availability and less interspecific competition. In contrast, natural environments are characterized by a random distribution of host plants, low plant density, high genetic diversity and uneven distribution of nutrients (Stukenbrock & McDonald, 2008). We should then expect phenotypic values of plants living in agroecosystems to be quite different from those measured in the wild. The ability of a pathogen to adapt to this new environment controlled by human therefore depends on its life history and potential for evolution of traits (McDonald & Linde, 2002; Giraud *et al.*, 2010). In a recent study, the term pestification was even introduced for the first time by Saleh *et al.* (2014) to describe the influence of rice domestication on its associated fungal pathogen *Magnaporthe oryzae*. In their study they revealed modifications of population genetic structure during the evolutionary history of the pathogen without any investigation on life history traits. Here we will define the term pestification as the process by which a pathogen originally living in the wild will become adapted to biotic and abiotic crop conditions. So what are the consequences of the transition of a wilderness characterized by a great heterogeneity of hosts to more homogeneous and dense agroecosystems? What are the impacts of domestication hosts on the evolution of history view associated pathogenic traits?



**Figure 1: Evolutionary consequences of (a) host domestication on phenotypic evolution on its pathogen (b) for two quantitative trait.**

The horizontal axes represent the biological parameters (component of phenotype) and the vertical axis represents fitness. The peak of distribution is the maximum fitness for the combination of two quantitative traits (adaptive peak). During domestication, a combination of traits is preferred and selected (a). The domestication of host has led to a new distinct peak, that defines the so-called "domestication syndrome" traits (green arrow). Two hypotheses can be made on the impact of domestication on the life history traits of the pathogen (b). In the first hypothesis, it is assumed that during the domestication process, the pathogen and host have co-evolved together. A new distinct peak appears (black star). The phenotypic value in this case is closely related to the fitness landscape of the host. In the second hypothesis, the white star symbolizes a new adaptive peak, which corresponding to a portion of phenotypic diversity present before domestication.

Considering wild and agricultural environments, several hypotheses on the impact of domestication on the life history traits of the pathogen are possible. Let consider that domestication of host has led to a new distinct peak, that defines the so-called "domestication syndrome" traits (Fig. 1a). On the one hand, we assume a low phenotypic plasticity in pathogen with a fitness landscape closely linked to the host's one. Assuming that pathogen

and host have co-evolved together, wild and domestic populations of the pathogen then have two separate fitness distributions (Fig. 1b – Hypothesis n°1). On the other hand, the change of adaptive peak in host does not necessarily lead to a parallel change in pathogen. Phenotypic changes in the hosts may have selected a part of the standing phenotypic diversity in the pathogen, leading to a new adaptive peak within the fitness distribution of wild phenotypes (Fig. 1b – Hypothesis n°2).

While life history traits selected during plant domestication are well known (for instance seed characteristics, flowering time, dormancy reduction), we still have no idea about the nature of the life history traits concomitantly selected in fungal pathogens. Among all, two categories are particularly relevant in epidemiology : growth and dispersal (Pringle & Taylor, 2002). First, intrinsic mycelial growth is a classical measurement of fitness (Brasier, 1999; Pringle & Taylor, 2002). Secondly, fungal pathogens are often characterized by a dispersive phase mediated by asexual (mitotic) spores also called conidia. Criteria like the number of conidia produced, their size, and their germination rate are good predictor of fitness (Pringle & Taylor, 2002) and should be relevant to determine the ability for dispersal. Deciphering how growth and dispersal of fungal pathogens have evolved during the domestication of their host would help us to better understand their adaptive dynamics.

The pathosystem *Malus spp.-Venturia inaequalis* is particularly relevant to investigate the changes in life history traits of pathogens adapting to domesticated hosts and agro-ecosystems. First, the life history of *V. inaequalis* confers to this ascomycete a significant potential for evolution (McDonald & Linde, 2002). As its reproduction occurs between strains of opposite mating types that infected the same leaf, this trait may facilitate prezygotic isolation process and therefore ecological divergence (Giraud *et al.*, 2010. Gladieux *et al.*, 2011). *V. inaequalis* produce both ascospores, spores as the outcome of meiosis, and conidia. Intrinsic dispersal ability of these two kinds of propagules is low, the most favorable way for the pathogen to achieve long distance dispersal is linked to human activities: the transport of fruits and infected plants (McHardy, 1996; Holb *et al.*, 2004). Actually, population structure of the pathogen is assumed to be linked with the history of the spread of its host (Wirth *et al.*, 2005; Gladieux *et al.*, 2008). The cultivated apple was domesticated in Central Asia from *Malus sieversii* (Velasco *et al.*, 2010; Cornille *et al.*, 2012 ; 2014). This species is located in the Tien Shan mountains (border of Kazakhstan, Kirghizstan and Xinjiang Province in China) to the Caspian Sea (Morgan & Richards, 1993) and forms large forests (Harris *et al.*, 2002). The migration westwards have involved contact and hybridization with closest wild relatives

growing along the Silk Road, *M. sylvestris* in Europe, *M. orientalis* in the Caucasus (Cornille *et al.*, 2013). These wild apples have then also contributed to the genetic mixing domesticated apple. Third, previous genetic studies on *V. inaequalis* give us many clues about pathogen origin. A global analysis of the diversity of *V. inaequalis* first showed that the pathogen shared a common origin with its host in Central Asia (Gladieux *et al.*, 2008). Impact of apple domestication on genetic structure of the pathogen has then been investigated (Gladieux *et al.*, 2010). Focusing their study in Central Asia they identified a population sampled on *M. sieversii* in natural forests of Kazakhstan that they considered to be a relic of the ancestral population from which have been derived populations currently present in agroecosystems worldwide. More recently, a study focusing on pathogenicity traits of this ancestral population have shown that apple domestication had markedly modified pathogenicity of *V. inaequalis* (Lê Van *et al.*, 2012). Indeed, domestication process was associated with the acquisition of virulence and increase of aggressiveness in *V. inaequalis*. Thus availability of valuable genetic resources and knowledge on the pathosystem provide an excellent framework to investigate impact of host domestication on evolution of pathogen life history traits.

Here, we compared several life history traits of the pathogen sampled on wild apple trees *Malus sieversii* within the ancestral population and within the domestic one in Kazakhstan. Previous studies have shown that *Venturia inaequalis* strains found in the Central Asia plains share the same characteristics whatever they are sampled in *Malus x domestica* or *M. sieversii* (Gladieux *et al.*, 2010; Lê Van *et al.*, 2012). We choose traits known in pathogens to be involved in dispersal and colonization success: mycelium growth, asexual spore germination rate, number of asexual spores and size of asexual spores. We address the following questions: (i) Are strains sampled in domestic habitat phenotypically different from those coming from wild habitats? If so, which traits are involved in these differences? (ii) Did domestication of apple trees drive evolution of *V. inaequalis* to a new adaptive peak rather than to a standing phenotypic combination selected for life into agroecosystems? (iii) At last, what are implications of an eventual phenotypic change for the life history of *V. inaequalis* in agroecosystems?

## MATERIALS & METHODS

We state that no permit was necessary to sample apple tree leaves containing the fungus *Venturia inaequalis* in the two locations CAM and CAP in Kazakhstan. The field study did not involve endangered nor protected species.

### Fungal isolates

This study was based on a total of 60 isolates of *Venturia inaequalis* sampled on the wild apple *M. sieversii* in Kazakhstan, 30 were collected in the wild forests area (43.23°N, 77.28°E) and 30 in the suburbs or agricultural area around Almaty (43.25°N, 76.91°E), the former capital of Kazakhstan. All these isolates were collected in 2006. Throughout the study, the terms wild population and domestic population are used to make reference to a collection of individuals sampled in non-anthropic area and in anthropic area respectively as defined in Gladieux *et al.* (2010)

### Phenotypic traits measurements

#### Mycelial Growth

Estimation of the mycelium growth capacity of isolates was performed on malt agar Petri dishes. For each isolate, eight plugs of mycelium from a 2-week-old culture were placed on a malt Petri dish. Cultures were then kept at 18 °C. Considering that *V. inaequalis* growth is circular, we measure mycelium growth diameter using a ruler 14 days after plugs deposit. Two measurements were performed for each agar plug, which represents 16 measurements per isolate.

#### Spore traits measurements

Spore traits analysis included the measures of the spore germination rate, sporulation capacity and spore sizes. Initial inocula used were obtained by growing monoconidial isolates of the fungus on cellophane sheets deposited onto malt agar medium (Bus *et al.* 2005).

Germination rates were assessed for each monoconidial suspension on malt agar plates. Petri dishes of malt agar covered with a cellophane sheet were kept in the darkness at 18 °C to allow conidia germination. After 24 hours, the number of spores germinated and non-germinated were counted using a microscope. Germination rate *in vitro* was then defined as



the percentage of germinated spores after 24 hours. For each isolate, one germination rate was calculated.

Concerning sporulation capacity and spores size measurements, 10 µl of the initial inoculum concentrated at  $10^5$  conidia ml<sup>-1</sup> were deposited on 4 four Petri dishes of malt agar covered by cellophane sheets calibrated to 3 cm<sup>2</sup>. Petri dishes were then stored for 7 days at 18 °C under the light. After this period, cellophanes were water rinsed for measurements of spore sizes and spores concentration using a spore counter (Counter Multisizer, Beckman). Sixteen measurements were made for each isolate. The counter was calibrated to count only particles whose diameter was in the range from 6.542 to 12.57 µm (range size of *Venturia inaequalis* conidia).

### **Statistical analyses of the phenotypic data**

All statistical analyses were performed using R Version 3.0.0 (R-Development-Core-Team, 2013). In order to test for independence between life history traits, we performed a Principal Component Analysis (PCA) using the “FactoMineR” package (Husson *et al.*, 2014) on all the dataset. We also used pairwise Pearson's correlation tests between traits within the two populations in order to test any trade-off change within each population. Pearson's correlation coefficients were estimated using the “corrgram” package (Wright, 2013).

For each life history trait, distributions were checked for normality using Kolmogorov-Smirnov test. To test for differences between populations in each phenotypic variable we used different procedure according to the availability of replicates. For differences between populations in sporulation capacity and spore size, an analysis of variance (ANOVA) was carried with a linear mixed-effect model (LME) implemented in the “nlme” package (Pinheiro *et al.*, 2014). The « individual » factor was treated as random effect and nested in « population ». Germination rate for each isolate were analyzed using student test. Percentages have been transformed by a logit transformation. Population effect on variation in mycelial growth was analyzed using an ANOVA. In all ANOVA, normality of residuals was checked.

## RESULTS

### Quantitative distributions of life history trait variables

Analyses of the mycelial growth were based on diameter measurements 24 days after the agar deposits. Overall mycelial growth was between 1.29 et 3.53 cm : data ranging from 1.28 to 2.94 cm (mean $\pm$ SD: 2,22  $\pm$  0.41) in the wild population and from 1.39 to 3.53 cm (mean $\pm$ SD: 2.14 $\pm$  0.50) in the domestic population. Spores analysis included comparison of the spore germination rates, sporulation capacities and spore sizes.

For the overall sample, germination rate varied between 45 and 96% : the rate ranged from 54 to 96% (mean $\pm$ SD: 81  $\pm$  10.82) in the wild population, and from 45 to 96% (mean $\pm$ SD: 79.6  $\pm$  14.08) in the domestic one. Sporulation capacity fluctuated much more than the other characters. The lowest rate of sporulation (3 250 spores ml<sup>-1</sup>) was observed in a strain of the wild population and the highest in a strain originating from the domestic population 534 400.

Distributions for each population were from 3 250 to 298 350 spores ml<sup>-1</sup> (mean $\pm$ SD: 122 400.8  $\pm$  80 008.83) and from 39 600 to 534 400 spores ml<sup>-1</sup> (mean $\pm$ SD: 199 955  $\pm$  108 084.1 ) for wild and domestic populations, respectively. Concerning the spores size, isolates from the wild population had sizes ranging from 8.12 to 10.15  $\mu$ m (mean $\pm$ SD: 9.26  $\pm$  0.44), while those from domestic isolates were between 8.45 to 10.45  $\mu$ m (mean $\pm$ SD: 9.55  $\pm$  0.36).

### Correlation between life history traits

Pearson's correlations were tested between life history traits and correlation analysis were performed pairwise. Spore sizes were significantly correlated with the sporulation capacity in each population (Table 1). No other correlation with mycelial growth and germination rate was detected. Germination rates of conidia were not correlated with any other trait, which means that the ability to germinate does not determine the size of the spores, nor the sporulation capacity.

The spore sizes evaluated were significantly correlated with sporulation capacity, with Pearson's correlation coefficients of R = 0.64 in the wild population (P<0.0001) and R = 0.48 in the domestic one (P<0.0076) (Table 1).

**Table 1 : Pearson's correlation coefficients between life history traits for the wild population (a) and domestic population (b)**

Pearson's correlation coefficients between life history traits (lower part) and significant value (upper part) for wild population (a) and domestic population (b). Pearson's correlation coefficients were calculated from averaged data from each strain for the sporulation capacity, spore size and mycelial growth. Statistically significant correlation coefficients ( $P < 0.05$ ) are in bold.

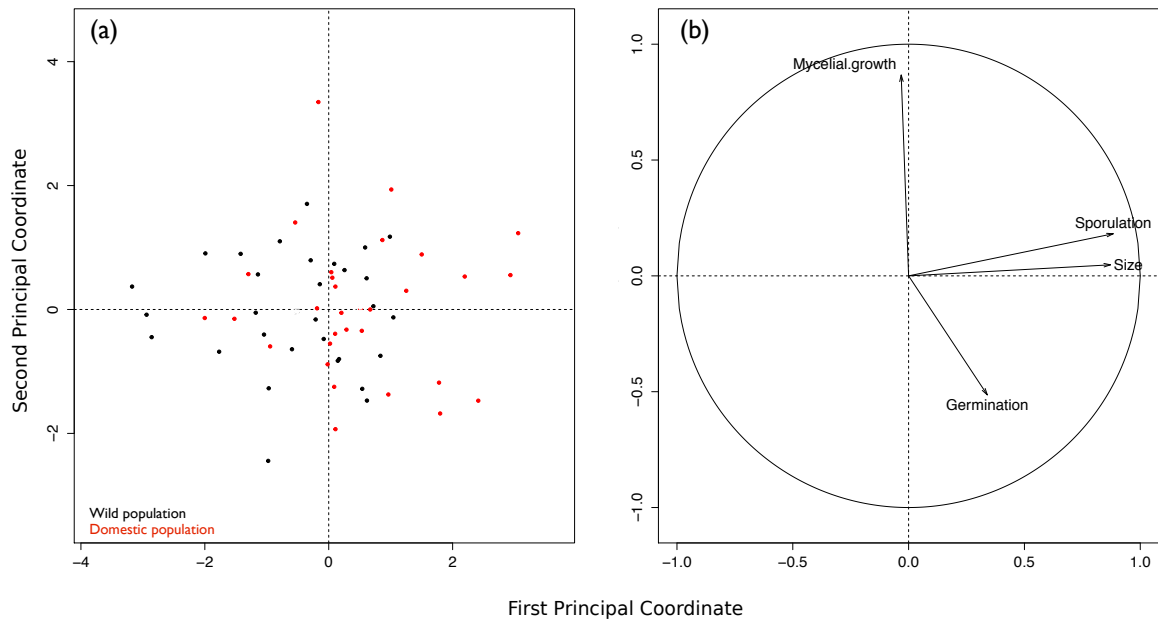
\* $P < 0.05$

\*\* $P < 0.001$

(a)	Mycelial growth	Sporulation capacity	Spore size	Germination rate
<b>Mycelial growth</b>		0.8391	0.7994	0.7785
<b>Sporulation capacity</b>	-0.0400		0.0001**	0.8598
<b>Spore size</b>	-0.0500	<b>0.6400</b>		0.9536
<b>Germination rate</b>	-0.0600	0.0300	0.0100	

(b)	Mycelial growth	Sporulation capacity	Spore size	Germination rate
<b>Mycelial growth</b>		0.2716	0.8641	0.6634
<b>Sporulation capacity</b>	0.2100		0.0076*	0.1287
<b>Spore size</b>	-0.0300	<b>0.4800</b>		0.2147
<b>Germination rate</b>	-0.0800	0.2800	0.2300	

To further characterize correlations, a PCA (Principal Component Analysis) was performed on all life history traits (Fig. 2). Our results clearly show that spore production and spore size were orthogonal with mycelial growth. The growth was independent of ability to produce asexual spores. This result confirms that sporulation capacity and spore size were not correlated to mycelial growth and germination rate. In fact variances in both spore size and sporulation were supported by the first axis, which 46.35% of the overall variance.



**Figure 2 : Principal coordinate analysis among 60 samples of *Venturia inaequalis*.**

The first and second principal coordinates components explain 40.39% and 35.48 % of the overall variation, respectively. This representation a) shows the plot of individuals on the two first principal components. Strains from domestic and wild habitats are separated according to their sampling site. The correlation circle b) shows the correlations between variables. It indicates the orthogonality between mycelial growth and germination rate on the one hand and sporulation capacity and spore size on the other hand. Variation explained by the first axis is mainly due to spore size and sporulation (b). These two variables should then explain the differences between wild and domestic strains as shown in (a).

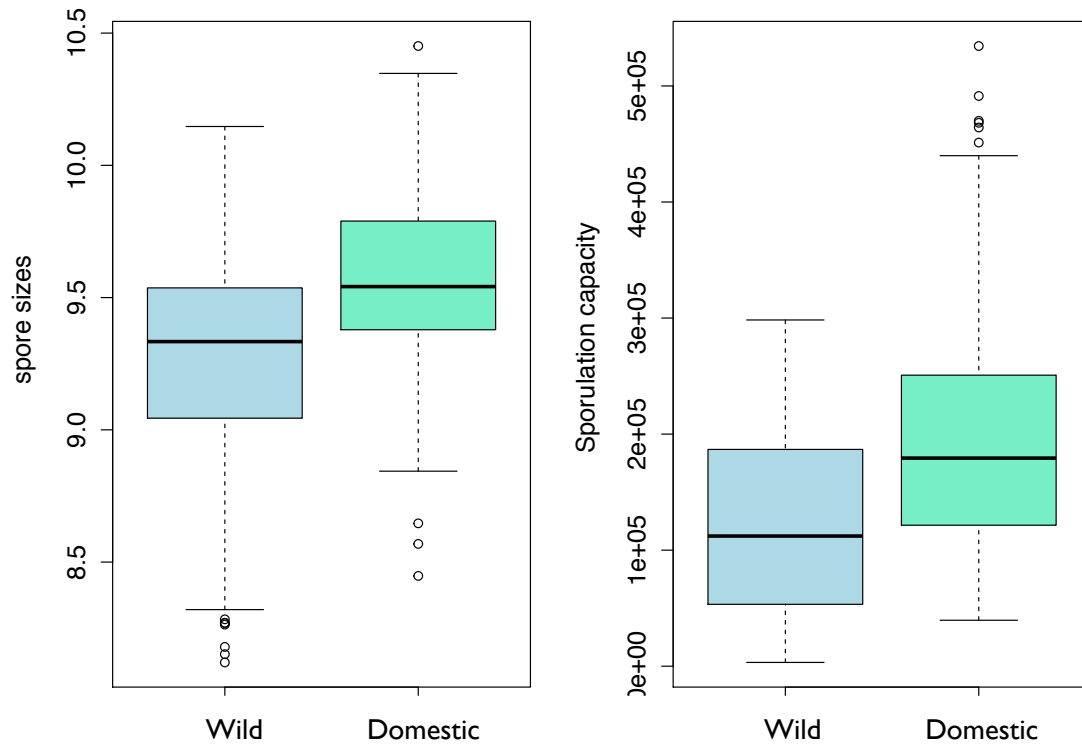
Life traits differences between the wild and domestic populations

Germination rate of the wild population was not significantly different from the rate of the domestic one, it was 81 % and 79.6%, respectively (Student test ;  $t=0.29$  ;  $P=0.77$ ).

Mycelial growth of the wild population was not significantly different from the growth of domestic population. 24 days after deposition, average mycelial growth was 2.21 and 2.13 for the wild and domestic populations respectively (ANOVA,  $F=0.40$  ;  $P=0.52$ ).

A significant difference of sporulation capacity between the two populations was detected (ANOVA ;  $F=188$  ;  $P<0.0001$ ). The domestic population had a sporulation rate higher than that of wild population. The average rate in the domestic population was  $199\,955\text{ spores ml}^{-1}$ , while  $122\,400\text{ spores ml}^{-1}$  in the wild population (Fig. 3). More, spore size of isolates from the wild population was significantly smaller than size spore of isolates from the domestic population (ANOVA ;  $F=19085$  ;  $P<0.0001$ ). Average size of the domestic population was  $9.5\text{ }\mu\text{m}$ , whereas this of the wild one was  $9.2\text{ }\mu\text{m}$  (Fig. 3). For these two phenotypic analyses, we noted that the distributions of the values of the two populations had a large area of overlapping. Almost all of the distribution of values of wild isolates was encompassed by the distribution of the values of domestic isolates. This result suggests that isolates sampled in the anthropic area were a sampling of the largest and most sporulating isolates of the wild population (Fig. 3).

Our results demonstrated that apple domestication promoted modifications in traits related to dispersal in the agent of the apple scab, *Venturia inaequalis*. Among the four life history traits analyzed, both spore production and size showed significant difference between the two sampling sites tested. Our results show that diversity for spore size and production is significantly higher in wild habitats than in agricultural ones.



**Figure 3 : Average spore sizes and sporulation capacity between wild population (blue boxplot) and domestic population (green boxplot).**

The horizontal line in the boxplots represents the median. Sporulation average in wild population: 122 400 spores ml<sup>-1</sup> and in domestic population = 199 955 spores ml<sup>-1</sup>. Spore size average in wild population = 9.2 µm and in domestic population = 9.5 µm.

## DISCUSSION

Impact of plant domestication on pathogen organisms has rarely been investigated although it is of major importance to understand consequences of modern human activities on disease emergence. Previous identification of fungal genetic resources that did not experience agriculture among all other human activities allow us to address this issue of great interest in evolutionary ecology. We used several phenotypic comparisons to investigate the changes of different life history traits involved in dispersal of the apple scab fungus, associated with the domestication of its host. We choose strains sampled on *M. sieversii*, the main wild asian ancestor of the domestic apple tree on an wild forest and on suburbs of Almaty. As previously mentioned, these last strains, although sampled on *M. sieversii*, have been shown to belong to the same populations as those living on domestic apple trees (Gladieux *et al.*, 2010; Lê Van *et al.*, 2012). Those strains were probably imported with domestic apple in Kazakhstan in the late XIX<sup>th</sup> century. Indeed as Kazakhstan was primarily inhabited by nomads, agriculture was not developed in this area, before the annexation by the Russian Empire and the colonization by Russian settlers. At last, choosing strains living on the same host species prevents from confounding effects due to adaptation to the genotype of hosts. Our two main findings are that (1) the transition from wild to agroecosystem was associated with an increase of the average of both spore sizes and sporulation capacity and (2) that distribution of each quantitative trait of the domestic population mostly overlapped with those of the wild population. Our results then suggest that transition from wild to orchard had considerable impact on fungal characters linked to its potential for dispersal by selection from standing phenotypic diversity. In the present study, we show that pestification of *V. inaequalis* in orchards led to an enhanced allocation in colonization ability for which potential capacity did already exist in the wild area.

### Correlation between fitness traits

The analysis of traits related to fitness is rendered complex as fitness is a synthetic function of both survival and reproductive success. Disentangling between these two components of fitness implies the measurement of several traits and the analysis of their pairwise correlations. It is worth considering what component of fitness is predicted by any given trait.

Our results clearly showed that spore production and spore size on the one hand were orthogonal with mycelial growth on the other hand (Fig. 2b). A likely explanation would be that, in *Venturia inaequalis*, growth is independent of ability to produce asexual spores. In other word the quantity of mycelium produced per time unit does not impact the net production of conidia. This result is at odds with observations made on *Aspergillus niger* where mycelial surface area was significantly correlated with spore number (de Visser *et al.*, 1997). Mycelium of *Venturia inaequalis* grows under the cuticle of the leaf and is therefore not visible. Aggressiveness is usually determined by the surface of symptom that is just the size of a lesion harboring conidiospores on the surface of the leaf. Our results would then suggest that this measurement of aggressiveness would not reflect intrinsic growth dynamics of *Venturia inaequalis* within the leaf. Nevertheless our results were obtained *in vitro* and then need to be validated *in situ*, for instance using thermography or fluorescence-based methods (Belin *et al.*, 2013). In addition, in our study, mycelial growth was analyzed from measurements of the diameter of mycelium surface area of each isolate. However, further analysis would be useful since the structure of the mycelium is formed of a succession of hyphae having similar branching pattern at different scales that fits a fractal geometry (Papagianni, 2006; Obert *et al.*, 1990). A description and modeling of fractal geometry mycelium would be more appropriate to evaluate mycelium growth. Spore size is known to be a trait in filamentous fungi related to dispersal abilities. Indeed, smaller fungal spores are currently considered as fitter than large spores as they are more effectively dispersed (Pringle & Taylor, 2002). Modeling of dynamics of airborne propagules clearly shows that spore size is negatively correlated with dispersability (Whitehead, 1969; Dix & Webster, 1995). However, in their recent theoretical study, looking at particles size ranging from 1 to 10µm (the actual size of conidia in our study), Norros *et al.* (2014) claim that "*the dispersiveness of especially the smallest spores is so high that landing may in fact be a bigger challenge for them than flying*". Conversely, larger spores are most likely to be deposited on hosts (Norros *et al.*, 2014; Meerts *et al.*, 1999; Dix & Webster, 1995). Thus, even if smaller spores disperse further than larger ones, the efficiency of dispersal in terms of probability of landing on a host favors larger spores.

Interestingly both spore production and size are strongly positively correlated in wild and agricultural habitats. Unlike plants and animals, here, the more propagules produced the bigger. This strong correlation probably reflects a strong linkage, even pleiotropy, between genes controlling spore production and spore size. Taking into account that larger spores



maximize their ability to be deposited on canopy (Norros *et al.*, 2014; Ingold, 1971), our results suggest that strains that produce the more asexual spores have the best efficiency of dispersal. Why did selection not favor strains with higher spore production and larger spores? The variability for both spore size and production in our dataset probably account for some sort of trade-off as previously mentioned by Meerts (1999) on basidiomycetes. Optimal spore sizes across species may vary according to life history. Thus as *Venturia inaequalis* disperses asexual spores on leaves and fruits mostly during splash events at short distance (Aylor 1998; McHardy *et al.*, 2001; Leroy *et al.*, 2013), we could then expect that larger spores will be favored in this species. We will develop this point further, at the comparison between strains from the wild and agricultural habitats.

At last, surprisingly, germination rates of conidia were not correlated with any other trait. Larger spores were expected to have greater germination rates (Pringle & Taylor, 2002) but to our knowledge there is no clear demonstration for fungi. Our results show that spore size does not determine the ability for germination. However, there is no data on survival rate of germinated spores. Indeed, observations shows that mortality can occur several hours to days after germinations. Such a correlation study between spore size and survival would deserve attention in the future.

#### Pathogen traits relative to efficiency of dispersal were impacted during wild to crop transition

Our results demonstrated that apple domestication promoted modifications in traits related to dispersal in the agent of the apple scab, *Venturia inaequalis*. Among the four life history traits analyzed, both spore production and size showed significant difference between the two sampling sites tested. Indeed, we observed an increase in the size and production of spores in population found within agroecosystems. This means that agroecosystems would have selected for a more efficient of short-distance dispersal. To understand this let compare wild habitats in Tien-Shan mountains of Kazakhstan with those found in suburbs of Almaty, the former capital of Kazakhstan. Wild apple trees, *M. sieversii*, from Tien-Shan mountains form dense and large forests extending along valleys but also fragmented habitats in meadows at the peripheral of forests. As expected for wild organisms, they exhibit high genetic diversity and then represent highly heterogeneous habitats for *Venturia inaequalis*. Conversely, *M. sieversii* found around Almaty grows in a landscape dominated by agriculture and especially apple orchards. As previously mentioned, orchards are characterized by a high homogeneity

of genotypes and density. First, our results show that diversity for spore size and production is significantly higher in wild habitats than in agricultural ones. This finding could reflect the greater heterogeneity of habitat. In fact, a strain adapted to its host would not be as fit in the neighboring host. Nevertheless, considering the potential for dispersal producing relatively smaller spores may ensure the ability to reach canopy or distant compatible habitats (Meerts, 1999; Norros *et al.*, 2014). In contrast, in habitats such as orchards, trees have the same or close genotypes and are nearby to each other. For a given *V. inaequalis* strain to invade, the best strategy would be to produce large conidia that would have better chances to be deposited on neighbor apple trees. Since the strain producing spores is already adapted to the genotype of its host, its asexual spores have a high likelihood to be also adapted to the hosts all around. Thus selecting for high sporulation would also select for high spores and potentially give an advantage in competition in an homogeneous habitat.

#### *Pestification by selection from standing phenotypic variation in the wild*

Plant and animals domestication process have often been associated to a strong disruptive selection favoring certain traits in each environment. In plants, transition between wild and cultivated forms was generally associated with different selection-targeted traits related to harvesting conditions, seed production or seedling competition (Harlan *et al.*, 1973).

Here we did not observe any disruptive selection in the different traits that we analyzed in *Venturia inaequalis*. Indeed, no actual *de novo* adaptation was observed in strains sampled in agricultural environment. Instead we observed that range size of each trait of the anthropized population mostly overlapped with those of the wild population. Indeed we showed that phenotypic diversity within the domestic population preexisted in the wild population at least on the traits we analyzed (Fig. 2a). This finding contrasts with description on many plants where domestication process creates a new adaptation niche, thus generating a strong disruptive selection of characters that is then associated to the “domestication syndrome”. There are some plausible explanations. The most trivial reason is that pathogen has not been selected by human for any agricultural or biotechnical trait of interest compared to plants or even domesticated fungi such as *Penicillium* (Ropars *et al.*, 2014) or *Saccharomyces* (Hittinger 2013). Second, as host tracking is a continuous process, we might thus hypothesis that plant modifications was progressive and so did not exert a strong selection contrary to host shift process. In the same line, it is to note that that *Malus x*

*domestica* remains genetically very close to its wild relative species *M. sieversii*, *M. sylvestris* and *M. orientalis* (Cornille *et al.*, 2012). Third, transition from wild to apple based-ecosystem may be too recent (*ca.* 4000-7000 years BC (Gladieux *et al.*, 2010) to expect a disruptive displacement of character by apparition of favorable mutations (but see Karasov *et al.*, 2010; Barton, 2010). Fourth, habitat of wild and cultivated apples did not really differ whatever its geographical distribution in the world, they clearly remain plants adapted to temperate climate. We might assume that this host trait has not imposed new dramatic abiotic conditions to the fungus. Similarly, a previous study on origin of *V. inaequalis* confirms that moving this fungus outside its native range had not induced shifts in its reproductive mode that is temperature dependent (Gladieux *et al.*, 2008).

We presented here one of the first study of evolution of fitness-related traits in a pathogenic fungus triggered by the domestication of its host. As previously introduced by Saleh *et al.* (2014) we called this process pestification. We found that this process led to a higher sporulation associated to a higher spore size. As these traits are highly correlated even in the wild we do not know which trait was selected for. However, selecting for those traits should increase the ability for efficient colonization in an homogeneous environment. In *Venturia inaequalis* pestification mostly proceeded by selection from standing phenotypic variation in the wild, rather than by progressive adaptation *de novo* as currently invoked in the literature on pathogen organisms. However, we do not know how much we can extend our conclusions to other pathogens as the evolution of fitness-related traits is highly dependent of the different life histories like the predominance of sexual over asexual reproduction (Pringle & Taylor, 2002). Nevertheless we found a high phenotypic diversity in the wild strains for the four traits we analyzed. We cannot exclude such a diversity for other traits like virulence or aggressiveness as suggested by Lê Van *et al.*, (2012). This study emphasizes the potential threat of populations of pathogens living in wild environment for durability of resistance in crops (Lê Van *et al.*, 2012; Monteil *et al.*, 2013; Leroy *et al.*, 2014).

## AKNOWLEDGMENTS

We would like to thank V. Caffier for all advices concerning experimental procedures. We also thank the french group "REID Champignons" for helpful comments. This work was granted by the SPE department of INRA and Region Pays de Loire and ANR-Bioadapt GANDALF (ANR-12-ADAP-0009).

## REFERENCES

- Abbo S, Pinhasi Van-Oss R, Gopher A, Saranga Y, Ofner I, Peleg Z. 2014.** Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in Plant Science* **19**: 351-360.
- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004.** Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**: 535-544.
- Aylor DE. 1998.** The aerobiology of apple scab. *Plant Disease* **82**: 838-848.
- Barton N. 2010.** Understanding adaptation in large populations. *PLoS Genetics* **6**: e1000987.
- Belin E, Rousseau D, Boureau T, Caffier V. 2013.** Thermography versus chlorophyll fluorescence imaging for detection and quantification of apple scab. *Computers and Electronics in Agriculture* **90**: 159-163.
- Brasier CM. 1999.** Fitness, continuous variation and selection in fungal populations: an ecological perspective. *Genetical Research* **77**: 213-218.
- Buckler ES, Thornsberry JM, Kresovich S. 2001.** Molecular diversity, structure and domestication of grasses. *Genetical Research* **77**: 213-218.
- Bus VG, Laurens FN, Van De Weg WE, Rusholme RL, Rikkerink EH, Gardiner SE, Bassett HC, Koddle LP, Plummer KM. 2005.** The Vh8 locus of a new gene-for-gene interaction between *Venturia inaequalis* and the wild apple *Malus sieversii* is closely linked to the Vh2 locus in *Malus pumila* R12740-7A. *New Phytologist* **166**: 1035-1049.
- Clement CR. 1999.** 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Economic Botany* **53**: 188-202.
- Cornille A, Giraud T, Bellard C, Tellier A, Le Cam B, Smulders MJ, Kleinchmit J, Roldan-Ruiz I, Gladieux P. 2013.** Postglacial recolonization history of the European crab apple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* **22**: 2249-2263.
- Cornille A, Giraud T, Smulders MJ, Roldan-Ruiz I, Gladieux P. 2014.** The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**: 57-65.

- Cornille A, Gladieux P, Smulders MJ, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L et al. 2012.** New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* **8**: e1002703.
- de Visser JAGM, Hoekstra RF, Van Den Ende V. 1997.** Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* **51**: 1499-1501.
- Dix NJ, Webster J. 1995.** *Fungal Ecology*. London, UK: Chapman et Hall.
- Gifford R. 2012.** Viral evolution in deep time : lentivirus and mammals. *Trends in Genetics* **28**: 89-100.
- Giraud T, Gladieux P, Gavrillets S. 2010.** Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution* **25**: 387-395.
- Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, Didelot F, Le Cam B. 2011.** Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology* **20**: 4521-4532.
- Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B. 2008.** On the Origin and Spread of the Scab Disease of Apple: Out of Central Asia. *PloS One* **1**: e1455.
- Gladieux P, Zhang XG, Roldan-Ruiz I, Caffier V, Leroy T, Devaux M, Van Glabeke S, Coart E, Le Cam B. 2010.** Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology* **19**: 658-674.
- Glémin S & Bataillon T. 2009.** A comparative view of the evolution of grasses under domestication. *New phytologist* **183**: 273-290.
- Hammer K. 1984.** Das Domestikations syndrom. *Kulturpflanze* **32**: 11-34.
- Harlan J, de Wet J, Price E. 1973.** Comparative evolution of cereals. *Evolution* **27**: 311-325.
- Harris SA, Robinson JP, Juniper BE. 2002.** Genetic clues to the origin of the apple. *Trends in Genetics* **18**: 426-430.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S et al. 2007.** Grinding up wheat: A massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution* **24**: 1506-1517.

- Hittinger CT. 2013.** Saccharomyces diversity and evolution: a budding model genus. *Trends in Genetics* **29**: 309-317.
- Holb I, Heijne B, Withagen J, Jeger M. 2004.** Dispersal of *Venturia inaequalis* ascospores and disease gradients from a defined Inoculum source. *Journal of Phytopathology* **152**: 639-646.
- Husson F, Josse J, Le S, Mazet J. 2013.** FactoMineR: multivariate exploratory data analysis and data mining with R. R package version 1.26.
- Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006.** Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences* **103**: 16666-16671.
- Ingold CT. 1971.** *Fungal spores: Their liberation and dispersal*. Oxford, UK: Clarendon press.
- Karasov T, Messer PW, Petrov DA. 2010.** Evidence that Adaptation in *Drosophila* Is Not Limited by Mutation at Single Sites. *PLoS Genetics* **6**: e1000924.
- Leroy T, Le Cam B, Lemaire C. 2014.** When virulence originates from non-agricultural hosts: New insights into plant breeding. *Infection, Genetics and Evolution*. doi:10.1016/j.meegid.2013.12.022.
- Leroy T, Lemaire C, Dunemann F, Le Cam B. 2013.** The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology* **13**: doi:10.1186/1471-2148-13-64.
- Lê Van A, Gladieux P, Lemaire C, Cornille A, Giraud T, Durel CE, Caffier V, Le Cam B. 2012.** Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications* **5**: 694-704.
- Li CB, Zhou AL, Sang T. 2006.** Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytologist* **170**: 185-193.
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V et al. 2009.** Population genomics of domestic and wild yeasts. *Nature* **458**: 337-341.
- Liu A, Burke JM. 2006.** Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* **173**: 321-330.

- McDonald BA, Linde C. 2002.** Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**: 349-379.
- McHardy WE, 1996.** *Apple Scab: Biology, Epidemiology and Management*. The American Phytopathological Society, USA: St Paul.
- McHardy WE, Gadoury DM, Gessler C. 2001.** Parasitic and biological fitness of *Venturia inaequalis*: Relationship to disease management strategies. *Plant Disease* **85**: 1036-1051.
- Meerts P. 1999.** The evolution of spores size in Agarics: do big mushrooms have big spores? *Journal of Evolutionary Biology* **12**: 161-165.
- Monteil CL, Cai R, Liu H, Mechan Llontop ME, Leman S, Studholme DJ, Morris CE, Vinatzer B.A. 2013.** Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytologist* **199**: 800-811.
- Morgan J, Richards A. 1993.** *The Book of Apples*. The Ebury Press, London.
- Norros V, Rannik U, Hussein T, Petäjä T, Vesala T, Ovaskainen O. 2014.** Do small spores disperse further than large spores? *Ecology* **95**: 1612-1621.
- Obert M, Pfeifer P, Sernetz M. 1990.** Microbial growth patterns described by fractal geometry. *Journal of Bacteriology* **172**: 1180–1185.
- Papagianni M. 2006.** Quantification of the fractal nature of mycelial aggregation in *Aspergillus niger* submerged cultures. *Microbial Cell Factories* **5**: doi:10.1186/1475-2859-5-5
- Pinheiro J, Bates D, Debroy S, Sarkar D. 2014.** *Nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-117.
- Pringle A, Taylor JW. 2002.** The fitness of filamentous fungi. *Trends in Microbiology* **10**: 474-481.
- R-Development-Core-Team. 2013.** R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing [<http://www.Rproject.org>].
- Ropars J, López-Villavicencio M, Dupont J, Snirc A, Gillot G, Coton M, Jany JL, Coton E, Giraud T. 2014.** Induction of sexual reproduction and genetic diversity in the cheese fungus *Penicillium roqueforti*. *Evolutionary Applications* **7**: 433-441.
- Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al.* 2010.** Whole-genome resequencing reveals loci under

selection during chicken domestication. *Nature* **464**: 587-591.

**Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D. 2014.** South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, *Magnaporthe oryzae*. *New Phytologist* **201**: 1440-1456.

**Stukenbrock EH, Mcdonald BA. 2008.** The origins of plant pathogens in agroecosystems. *Annual Review of Phytopathology* **46**: 75-100.

**Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004.** Selection versus demography: A multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution* **21**: 1214-1225.

**Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D et al. 2010.** The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**: 833-939.

**Wirth T, Meyer A, Achtman M. 2005.** Deciphering host migrations and origins by means of their microbes. *Molecular Ecology* **14**: 3289-3306.

**Wright K. 2013.** Corrgram: Calculates correlation of variables and displays the results graphically. R package version 1.5.

**Whitehead R. 1969.** Wind pollination in the Angiosperms: evolutionary and environmental considerations. *Evolution* **23**: 28-35.





## Chapitre 3

---

### « Signatures génomiques de l'impact de la domestication du pommier sur son pathogène *Venturia inaequalis* »

Dans le chapitre précédent, nous montrons que la domestication a provoqué des changements de traits d'histoire de vie chez *V. inaequalis*.

Le dernier objectif de cette thèse est d'identifier les signatures génomiques de l'adaptation de *V. inaequalis* aux agrosystèmes.

Cette étude se base sur le ré-séquencage du génome de 36 souches provenant de la zone de domestication du pommier. L'approche de génomique évolutive implique dans un premier temps de déterminer le scénario évolutif le plus probable (via une approche basée sur le spectre bidimensionnel de fréquence et par ABC). L'identification des régions génomiques touchées par la domestication du pommier s'effectuera selon l'histoire démographique inférée.



## **Genomic signatures of the impact of domestication of apple trees on it fungal pathogen *Venturia inaequalis*.**

**De Gracia M, Roux C, Fraïsse C, Bierne N, Sannier M, Le Cam B, Lemaire C.**

### **INTRODUCTION**

The ascomycete fungus *Venturia inaequalis* is the major pathogen of apple tree with a global distribution (Gladieux *et al.*, 2008). This pathogen originated in Central Asia and its historic worldwide spreading is assumed to be closely linked to that of its host (Gladieux *et al.*, 2008). The ancestral progenitor of the cultivated apple in Central Asia was *M. sieversii*, it still grows in the Tien Shan mountains (border of Kazakhstan, Kirghizstan and Xinjiang Province in China) to the Caspian Sea (Morgan & Richards, 1993) and forms large forests (Velasco *et al.*, 2010; Cornille *et al.*, 2012, Harris *et al.*, 2002). It is during the Neolithic period and especially during the development of the Silk Road that the seeds of wild apple trees were scattered, thus marking the beginning of apple cultures (Juniper & Mabberley, 2006; Wood, 2003; Harris *et al.*, 2002). Domesticated apple would then spread westward to Europe and eastward to China along the silk roads. Our previous studies showed that *V. inaequalis* spreading worldwide was associated to an host tracking process where pathogen had followed its host from Central Asia to orchards worldwide through the domestication process. When investigating *V. inaequalis* origin, we identified one genetically divergent population sampled on *M. sieversii* in natural forests of Kazakhstan that were named CAM population for Central Asia Mountains (Gladieux *et al.*, 2010). Collected in non anthropized areas this population characterized by the presence of rare alleles and high genetic diversity were then considered to be a relic of the ancestral population from which would have derived populations currently present in agroecosystems worldwide. More recently, a study focusing on history life traits of this ancestral population have shown that apple domestication had markedly modified its pathogenicity and its dispersion capacity (Lê Van *et al.*, 2012; De Gracia *et al.*, submitted to Plos One).

Examination of domesticated plants and their wild relatives indicates that human activities have strongly impacted their history life traits leading ultimately to larger fruits or naked grains. Although the genetic and phenotypic modifications of crops related to domestication have been extensively studied, evidences of phenotypic and genetic changes in their associated pathogens remain poorly understood. Except for domesticated fungal species as *Saccharomyces cerevisiae* (Gu *et al.* 2005) and *Aspergillus oryzae* (Rokas 2009), the sole study on impact of the plant domestication on pathogens changes is the wheat pathogen, *Mycosphaerella graminicola*. Using a genomic approach based on an NGS dataset and coalescent methods, Stukenbrock *et al.* (2011) showed that the process of domestication of wheat has exerted strong selection pressures on the fungus *M. graminicola* and then provided numerous candidate genes putatively involved in plant adaptation.

Recently, Leroy *et al.* (2013) underlined the importance to identify the more realistic evolutionary scenario of divergence between two populations to avoid misinterpretations of data generated by genome scans. According to Bierne *et al.* (2011), many genome scans studies aiming at identifying adaptive loci consecutive to a divergent process provide many false candidate genes because of inadequate knowledge of the demographic scenario. For instance loci displaying high genetic differentiation (estimated for instance by  $F_{st}$ ) are too often interpreted as loci involved in local adaptation to the habitat while confounding effects like genetic barriers involved in hybrid depression in secondary contacts (Bierne *et al.*, 2011), hierarchical population structure (Excoffier, 2009) or fractal landscape geometry (Fourcade *et al.*, 2013) may strongly bias interpretations about the actual nature of such loci. Until recently, the lack of suitable statistical methods for distinguishing between different demographic scenarios and estimating demographic and population parameters (Richards *et al.*, 2007) hindered the accurate and precise inference of complex scenarios. Approximate Bayesian computation (ABC) provides a robust framework, allowing more powerful inferences of a species' past demography (Beaumont *et al.*, 2002; Bertorelle *et al.*, 2010; Csilléry *et al.*, 2010). These methods have been quite widely applied in population genetics and have improved our vision of the different forces shaping genome diversification during speciation (Beaumont *et al.*, 2012; Marjoram & Tavaré, 2006; Hey & Pinho, 2012). According evolutionary model, simulations of a large number of datasets are performed to produce posterior distributions of the parameters that produced the observed dataset. The reliability of the procedure is then calculated using the Euclidean distance the summary statistics computed

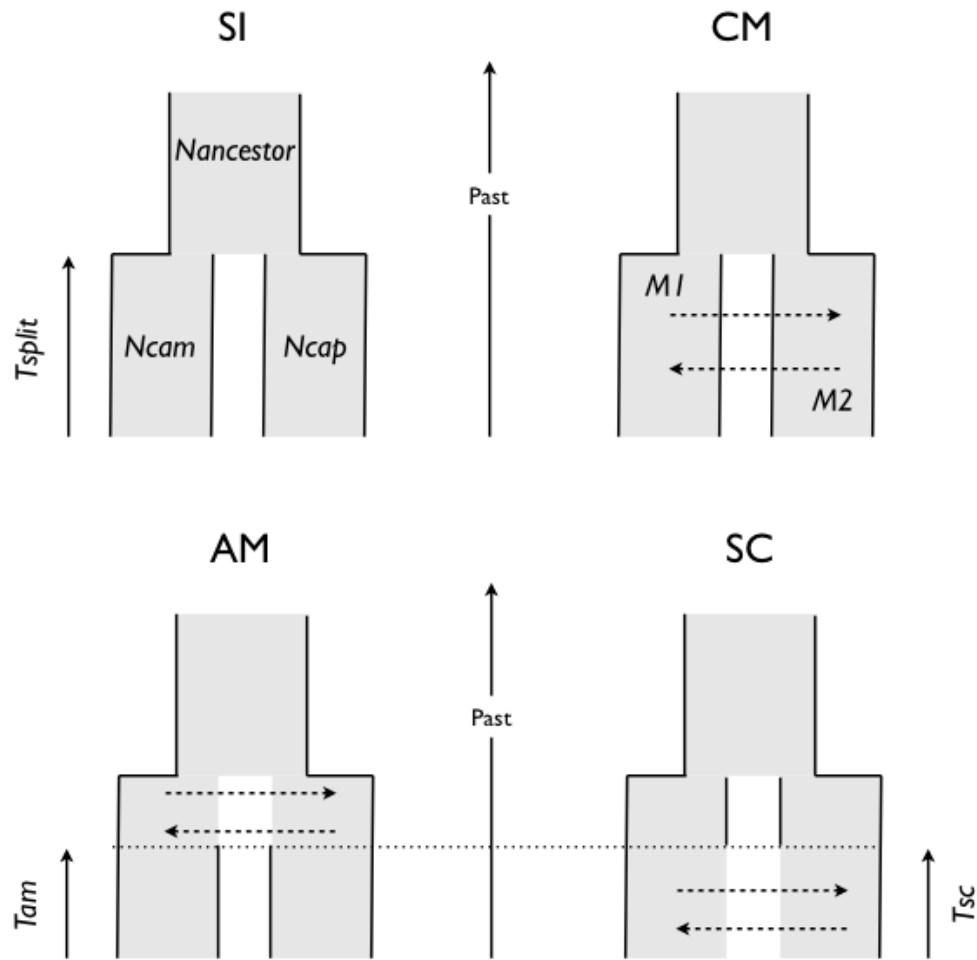
from the simulated datasets and those captured from real data (Beaumont *et al.*, 2002; Excoffier *et al.*, 2005). These methods allowed to compare different scenarios, to integrate different modes of introgression or to estimate divergence times, migration rates and effective populations sizes changes over time (Ross-Ibarra *et al.*, 2009; Gattepaille *et al.*, 2013) both in animals, plants or pathogens (Duvaux *et al.*, 2011; Roux *et al.*, 2010; Dutech *et al.*, 2012).

In the case of Kazakh populations of *V. inaequalis*, the host-tracking model was invoked to explain both genetic (Gladieux *et al.*, 2010) and phenotypic divergence (Lê Van *et al.*, 2012) between wild and agricultural strains. Such a model would leave peculiar genomic signatures. Host-tracking model would assume a recent divergence matching with the beginning of domestication of apples in Central Asia followed by gene flow between wild strains and those infecting new agricultural ancestors of *Malus x domestica*. This period of ancient gene flow would have stopped probably rather quickly, when apple cultures extended out of the range of the wild apple trees, *M. sieversii*. Wild and agricultural populations would have remained isolated until now. This scenario is adaptive as it assumes that the pathogen followed its host's changes by progressively accumulating small beneficial mutations. This model is also qualified as host-pathogen co-differentiation, as populations of pathogens differentiate from their wild relatives in response to differentiation of under domestication apple trees from the wild *M. sieversii*. Under this model we would expect a quick fixation of alleles by positive selection in the agricultural populations of *V. inaequalis*, leading to strong selective sweeps in the neighborhood of advantageous mutations. As divergence time between wild and agricultural populations is expected to be recent (<10,000 years that is less than 10,000 sexual generations) we then expect a high amount of shared polymorphisms between the two populations and incomplete lineage sorting. We also expect a low proportion of fixations within each populations, probably corresponding to loci involved in the adaptive process of host-tracking.

However, several alternative scenarios could be drawn. As previously proposed by Roux *et al.* (2011; 2012; 2014). In the first one, we considered that ancestral population has first split into two populations that have thereafter evolved in complete reproductive isolation without any gene flow (SI scenario named "Strict Isolation"). Under this scenario, the 2D-SFS should show high content of private derived alleles in each lineage. Under the second scenario, the gene flow is assumed throughout the divergence (scenario CM for "Constant Migration") giving a 2D-SFS a high amount of shared alleles between the two populations. AM scenario describes two currently isolated populations (AM for "Ancient Migration"), but having

experienced a period of flux gene in the early stages that followed their split. This scenario matches with the one of host-tracking, described above. The last demographic scenario describes a secondary contact (SC) followed by gene flow after a long period scalable without gene exchange. Under this last scenario, the 2D-SFS contains both shared and private alleles, depicting the complex situation where divergence would have promoted the occurrence of incompatibilities (endogenous selection) and local adaptation (exogenous selection) as reviewed in Bierne *et al.* (2011). Under the most likely scenario, we then characterized gene flow between the two populations and identified genes that have a strong genetic differentiation in order to address the question of the nature of reproductive barriers.

In a previous study we estimated divergence times among populations collected in anthropic and non anthropic zones using the the isolation-with-migration model implemented in the program IMA (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007) from a small microsatellite dataset. We then sequenced 36 genomes of *V. inaequalis* chosen among previously characterized CAM and CAP populations using NGS. Based on historic data of both apple and the agent of apple scab, we tested four evolutionary scenarios using polymorphism at 10435 coding sequences. The evolutionary genomic approach we apply here Approximate Bayesian Computation methods (ABC) (Beaumont, 2002) and an approach based on the two-dimensional site-frequency spectrum (2D-SFS) to (1) infer the most likely evolutionary scenario of the divergence between wild and domestic population, and (2) thereafter identify genomic regions of *V. inaequalis* impacted by apple trees domestication.



**Figure 1. Alternative scenario of divergence for CAM and CAP**

Fours classes of scenarios according to the pattern of migration : Strict Isolation (SI), Constant Migration (CM), Ancient Migration (AM) and Secondary Contact (SC).

For the last three types of scenarios (CM, AM and SC), three alternative models within each class of scenarios: homogeneous introgression (SI-Homo, AM-Homo, SC-Homo) and heterogeneous introgression (SI-Hetero, AM-Hetero, SC-Hetero).



## MATERIALS & METHODS

### Samples

We sampled 36 isolates of *Venturia inaequalis* on *Malus sieversii* trees in Kazakhstan in 2006. 17 isolates were collected in the wild forests in mountains and then defined as the CAM population (Central Asia Mountains). The others were collected in agricultural areas and defined as the CAP population (Central Asia Plain) (Gladieux *et al.*, 2011). Genomes sequence of *V. pirina* and *V. asperata* were used as outgroup (data unpublished)

### DNA extraction and sequencing

Genomic DNA of the 36 isolates was extracted from fresh mycelium according to Moller *et al.* (1992). For the library preparation, 15µg of DNA was sheared using the ultrasonicator Covaris E210. The sequencing library was prepared using the SPRIworks - Fragment Library System I (Beckman Coulter). Library selection was performed using a 2% agarose gel to recover fragments comprised between 250 and 350 bp. Several kits were used for sequencing: TruSeq Multiplex Sequencing Primer (Illumina), TruSeq PE (2x50bp). Each strain was independently sequenced. Considering the size of *V. inaequalis* genome that is around 73 Mb, sequencing was performed in order to obtain a coverage of 50X.

### Sequences Assembling – aligning genomes and identification of SNPs

Two algorithms were used to assemble genomes : SOAPdenovo2 (Zerbino *et al.*, 2008) and Velvet (Luo *et al.*, 2012) adapted for de novo genome assembly using next-generation sequencing (NGS) short reads. Genomes were aligned using the software Mugsy (Angiuoli & Salzberg, 2011). We only used the first part of the pipeline Mugsy, that is to say genome comparison two by two. Annotation of genomes and alignments performed with the software Mugsy. SNP prediction was performed using a software developed by the LIPM team at Toulouse : lipm\_maf2vcf.pl.

## Clustering methods and diversities analyses

Alignments of genes were analysed if the following criteria were gathered: possibility to align genes in at least of 10 isolates for each population and existence of an outgroup for each gene. Duplicates were removed. Classical population genetics indices like  $F_{st}$ , Tajima's  $D$ , Fay & Wu's  $H$  and Zeng's  $E$  statistics were estimated using Egglib python package (De Mita & Siol, 2012).

We used different methods to estimate the most likely number of clusters within the sampled populatoin. We performed a principal coordinate analysis (PCA), using ADEGENET (2013, R package version 1.4-2), generated with the R program Version 3.0.0 (R-Development-Core-Team, 2013). Identification of genetic structures in populations was based on the 1 794 853 SNP detected within the 36 isolates.

## Inferring ancestral demography

We used two different methods to retrace historic divergence of CAM and CAP populations

### a) Using the approximation diffusion implemented in $\partial a \partial i$

Two-dimensional site-frequency spectrum was estimated from the wild population and domestic population using a diffusion-based approach implemented in the program  $\partial a \partial i$ . Singletons were masked because their identification is prone to false positives. The joint frequency spectrum data were polarized using the *Venturia pirina* genome sequence (available in the laboratory). We used the models developed by C. Fraisse et al. (comm. pers). These models are related to the models SI, CM, AM and SC described above. As the SI model did not involve any gene flow, no other versions exists for this model. We also used IM, AM and SC models in their simple versions. For IM, AM and SC models we also designed more complex models including two migration rates (2M) corresponding to the migration rates of neutral loci and loci in genomic island respectively. We then call these models IM2M, AM2M and SC2M to take into account genome heterogeneity of introgression rates due to genetic barriers. To these models, we also can add parameters  $P1$  and  $P2$  giving the proportion of loci

evolving neutrally respectively in population 1 and 2. This parameter takes into account the occurrence of neutral loci linked to mutations involved in genomic islands. We called these models IM2M2P, AM2M2P and SC2M2P. For models AM and SC, we also assumed a new category of models assuming two periods of ancient migration for model AM or two periods of secondary contacts for SC model. Combining this to previously described model we then draw PAM and PSC models as well as PAM2M2P and PSC2M2P models. Another parameter noted O, permit to estimate the proportion of loci mis-oriented, that is loci for which the derived allele was not accurately defined from outgroup. We then compared these 14 models using likelihood ratio tests (LRT) between nested models. However, as the use of LRT assumes independence of SNP, we conducted these analyses using SNP distant of at least 5kb from each other, that is 2743 SNPs. Parameters were estimated from the best inferred model.

## **b) Using Approximate Bayesian Computation (ABC)**

### ***Coalescent simulations***

We compared four scenarios describing the divergence of two sister lineages: three scenarios allowing inter-specific effective migration (Isolation with Migration, IM; Ancient Migration, AM; and Secondary Contact, SC) and one without gene flow since the speciation event (Strict Isolation, SI). They all describe a sudden split at the time  $T_{\text{split}}$  of an ancestral population (of effective size  $N_a$ ) into two daughter populations of constant sizes ( $N_{\text{CAM}}$  and  $N_{\text{CAP}}$ ). The different scenarios differ by their temporal pattern of migration. In the AM scenario gene flow is restricted to the early stages of speciation, whereas in the SC scenario the two daughter populations begin to evolve in strict isolation forward in time, and then experience secondary contact. Finally, the IM model assumes continuous gene flow between populations. We used the scaled migration rates  $M=4Nm$  (with  $M_1$  the migration rate from *CAP* to *CAM* and  $M_2$  the migration rate into the *CAP* gene pool), where  $m$  is the fraction of the population made up of new migrants each generation. For the AM, SC and IM scenarios allowing gene flow, we simulated two alternative models with an introgression rate shared by all loci ("homogeneous" model) or varying throughout genome ("heterogeneous" model), making a total of seven compared models. Following Roux *et al.* (2014) the "heterogeneous" models consist of two groups of loci associated to different migration rates, with a proportion  $p$  of loci having a

reduced gene flow when compared to the other group of proportion  $(1-p)$ . For each model, five million of multilocus simulations were performed using large uniform prior distributions for all parameters, and identical for parameters shared among models. Prior distributions for  $\theta_{CAM}/\theta_{ref}$ ,  $\theta_{CAP}/\theta_{ref}$  and  $\theta_A/\theta_{ref}$  were uniform on the interval 0-40 with  $\theta_{ref}=4.N_{ref}.\mu$ .  $N_{ref}$  is the effective number of individuals of a reference population used in coalescent simulations, arbitrarily fixed at 100,000, and  $\mu$  the mutation rate of  $5 \times 10^{-8}$ /bp/generation. We sampled  $T_{split}/4.N_{ref}$  from the interval 0-10,  $0-4 \times 10^6$  generations in demographic units. The parameters  $T_{iso}$  and  $T_{SC}$  were drawn from a uniform distribution on the interval  $0-T_{split}$ . In the “homogeneous” models, values of the two migration rate parameters  $M_1$  and  $M_2$  were randomly sampled from the uniform prior interval 0-40 for all loci. Prior distributions were computed using a modified version of the PriorGen software, and coalescent simulations were run using MsnSam (Ross-Ibarra *et al.*, 2008), a modified version of the ms program (Hudson, 2002) that allows for different sample sizes at each locus under an infinite site mutation model.

### ***Model testing***

We compared summary statistics of the observed datasets to five million of random simulations under each scenario. Posterior probabilities for each candidate model were approximated using a feed-forward neural network implementing a non-linear multivariate regression by considering the model itself as an additional parameter to be inferred under the ABC framework using the R package “abc” (Csillery *et al.*, 2012). The 7,000 replicated simulations nearest to the observed values for the summary statistics were selected, and these were weighted by an Epanechnikov kernel that reaches a maximum when  $S_{obs}=S_{sim}$ . Computations were performed using 50 trained neural networks and 15 hidden networks in the regression.

### ***Parameter estimation***

We first estimated parameters of the best support SC\_Homo scenario using the same ABC framework. Parameters were first log-tangent transformed (Hamilton *et al.*, 2005), which

allows minimising heteroscedasticity problems during regressions. Transformed and adjusted parameter values were then back-transformed to obtain the posteriors. Only the 1,000 replicate simulations with the smallest associated Euclidean distance  $\delta = ||S_{\text{obs}} - S_{\text{sim}}||$  were considered. The joint posterior distribution of parameters describing the best model was then obtained by weighted non-linear multivariate regressions of the parameters on the summary-statistics (Blum & François, 2009). For each regression, 50 feed-forward neural networks and 15 hidden networks were trained using the R package “abc” (Csillery *et al.*, 2012).

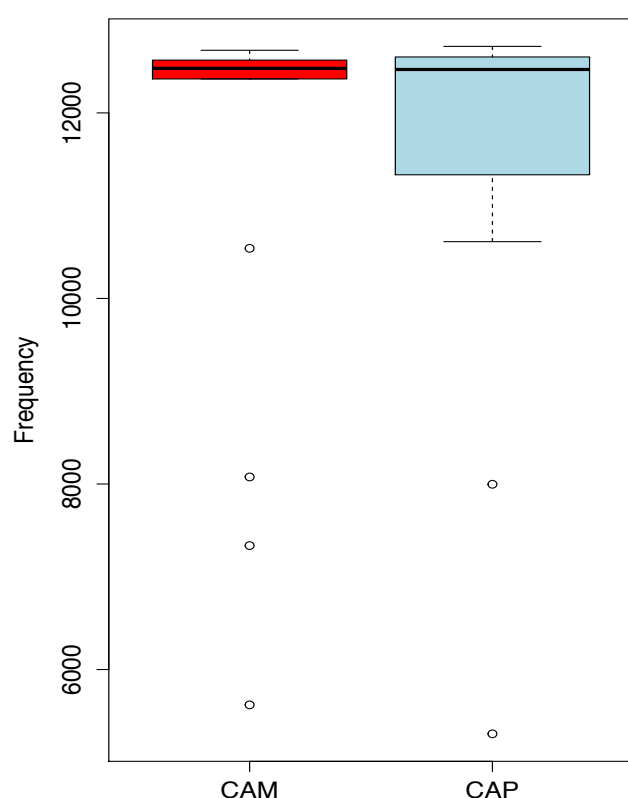
### **Functional annotation**

From results obtained with Egglib, an alignment of a genes set whose  $F_{st}$  between the two populations was greater than 0.7 was performed using Blast2GO. Blast2GO (Conesa *et al.*, 2005) is a tool for functional annotation of sequences and the analysis of annotation data. The method uses local sequence alignments (BLAST) to find similar sequences (potential homologs) for one or several input sequences. The program extracts all GO terms associated to each of the obtained hits and returns an evaluated GO annotation for the query sequence.

## RESULTS

### Datasets

Informations about the genome assembled using Velvet and SOAPdenovo2 software are shown in the table I. The average genome size assembled in the CAM population (mean±SD [standard deviation] : 40 948 610 pb ±4 691 682) is not significantly different from that of the CAP population (mean±SD : 38 844 256 pb ± 2 339 511) (Wilcoxon signed-rank test,  $W = 227$ ,  $Pvalue = 0.03833$ ). However, we can see that the sequenced genomes in the area of apple domestication have an average size of 2 times smaller than the reference strain genome EUB04 (73 Mb).



Annotation of the reference strain EUB04 detected 15,319 genes. In this study 14 078 genes alignments were performed that means that 1,241 genes were not detected in Kazakh strains. Distribution of gene sizes ranged from 5,619 to 12,675 genes (mean±SD: 11,423.18 ± 2,205.432) and from 5,307 to 12,717 genes (mean±SD: 11,567.74 ± 1,905.984) for CAM and CAP respectively (Figure 2). Numbers of genes of the CAM population were not significantly different from the CAP population (Wilcoxon signed-rank test,  $W = 150$ ,  $Pvalue = 0.7306$ ).

**Figure 2 . Average number of genes per population**

CAM (Central Asia Mountains) : wild population

CAP (Central Asia Plains) : domestic population

The 14,078 alignments were filtered. Finally, the diversity analysis was based on 10,994 alignments. Among the 10,994 alignments, 4,205 included an outgroup (*V. pirina*). 1,794,853 SNPs were detected.

Table I: Composition of genomes assembled

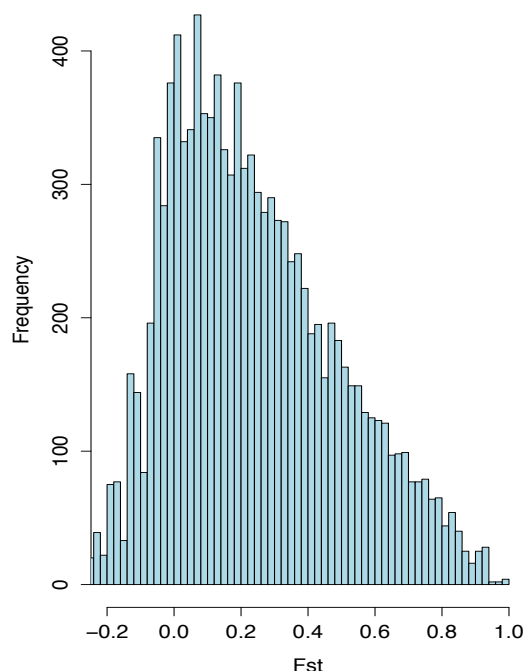
The table summarizes the data assembly for the 36 strains of the study (CAM population : red / CAP population : blue). The N50 length is defined as the length for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs,

	2446	2447	2222	2450	2451	2453	2454	2455	2456	2457	2458	2225	2459	2226	2460	2461
Assembly size	45739952	43010849	33890745	44498662	43584037	47607347	45462274	41045789	42641730	30751836	35972871	44214837	35214675	41291663	41485077	37311094
Nb	11296	10153	5014	11221	8996	8147	11796	8652	9833	9352	6074	10357	6053	6764	8266	5172
N50 size	39961	42783	18063	44124	44423	42632	44321	42242	42206	6412	29827	41989	18494	44022	43583	36534
Nb	329	285	555	285	287	332	306	287	290	1463	356	310	580	290	276	302

	2462	2463	2227	2228	2464	2466	2467	2468	2469	2470	2471	2472	2229	2473	2474	2475	2476	2230	2232
Assembly size	36722702	41850879	37311004	36351072	36383275	39792509	40439174	41989459	35466717	39634247	36305087	38531270	40837719	35088623	40301462	42813215	39631570	40133764	38470278
Nb	6223	8954	8047	4327	4424	6496	7132	7614	7417	6555	4512	5888	8032	4192	8185	9558	6859	8367	662
N50 size	31036	49109	22347	38586	40431	53837	47730	53550	12507	57093	35281	55445	50643	29779	55813	54582	51184	42892	48981
Nb	365	240	502	276	278	216	242	231	834	209	289	209	247	349	208	227	237	284	232

### Polymorphism and clustering method

To evaluate the demography and historical context of *V. inaequalis* divergence in its center of diversity, we estimated levels of the nucleotide diversity in the genomes of 36 isolates

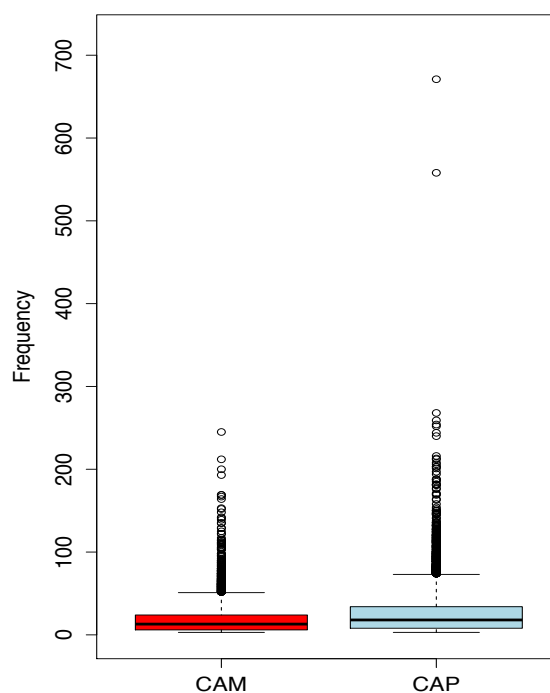


**Figure 3 . Genetic differentiation between CAM and CAP**

We estimated the number of polymorphisms sites. A significant difference was observed between the number of polymorphic sites in the CAM population (mean $\pm$ SD : 9.0714  $\pm$  15.7537) and CAP population (mean $\pm$ SD : 22.5845  $\pm$  25.9916) (Wilcoxon signed- rank test,  $V = 2\,890\,529$ ,  $p\text{-value} < 2.2 \times 10^{-16}$ ). (Figure 4). CAP population was more polymorphic than the CAM population.

originating from wild and domestic populations.

We analyzed genetic differentiation between the two populations on 10,994 gene alignments (Figure 3). The calculation of  $F_{st}$  was realized with the Egglib software. Differentiation indices revealed a considerable heterogeneity.  $F_{st}$  ranged from 0.4430 to 1. The average value was 0.2409 (SD : 0.2494). A  $F_{st}$  of 1 indicates a total differentiation at this locus. Based on this analysis, we obtained more than 602 genes whose alignments  $F_{st}$  were higher than 0.7, that represents 18.26 % of the dataset.

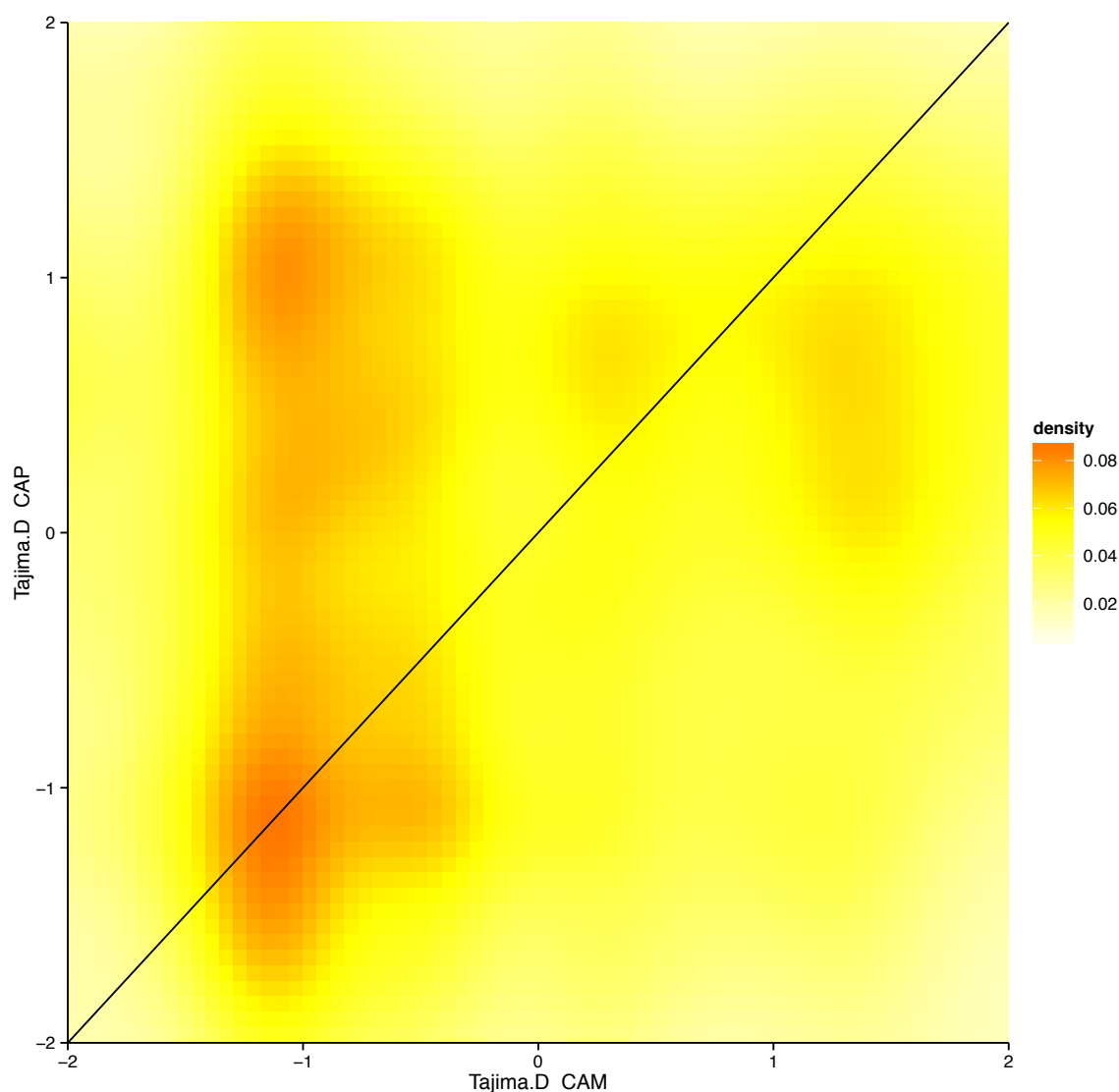


**Figure 4 . Number of polymorphic sites in the CAM and CAP populations**



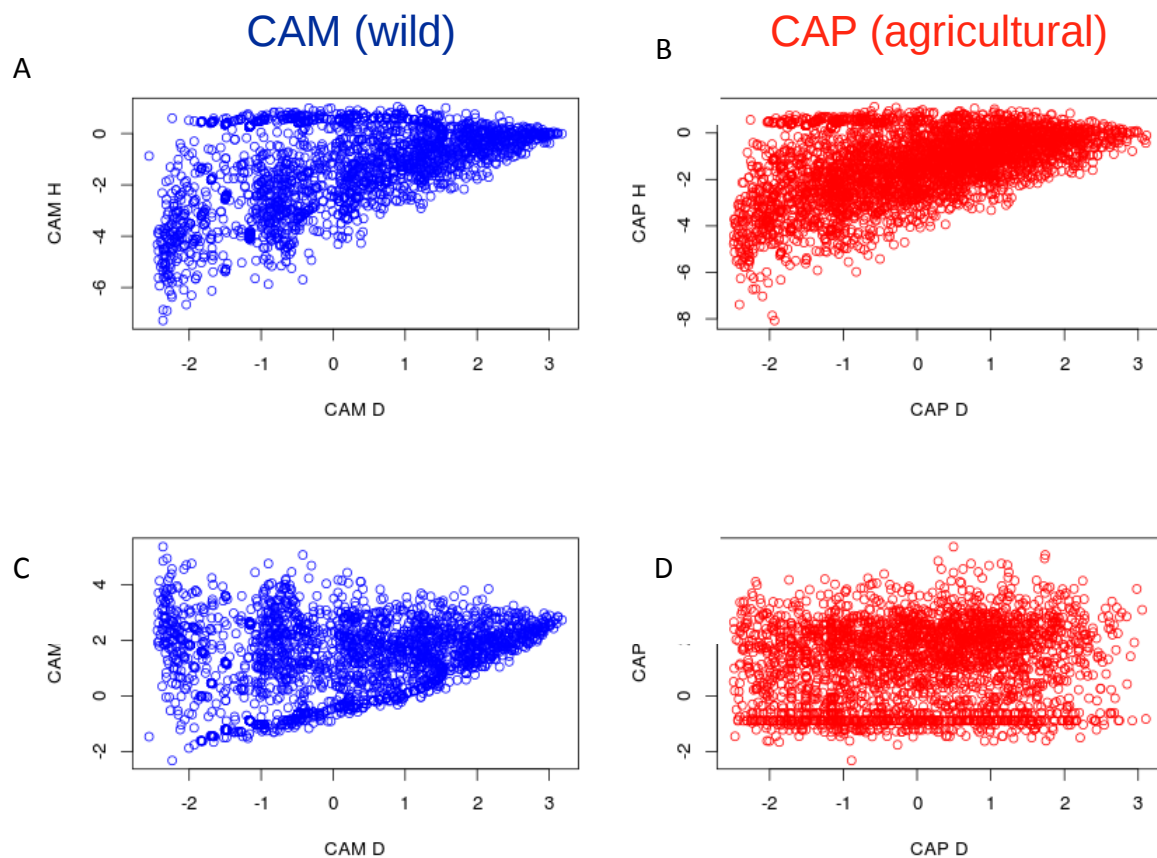
Tajima's D multilocus analysis showed no difference within the CAM population (mean $\pm$ SD : 0.0467  $\pm$  1.3990) and within the CAP population (mean $\pm$ SD : 0.0129 $\pm$  1.2685) (Wilcoxon signed-rank test, W = 43 326 877, Pvalue=0.4153).

Figure 5 represents the density map D Tajima. Each point corresponds to the intersection of Tajima's D between CAM and CAP for one gene. The more red area the higher the density of this area is strong and representative. This density map showed that Tajima's D varies around -1 in CAM and between -2 and 1 in CAP. Negative values reflect an excess of rare alleles that may be due to a demographic expansion while positive values indicate the action population decline. Hypotheses invoking selection can be made for individual loci but not for general behavior of thousands of loci.



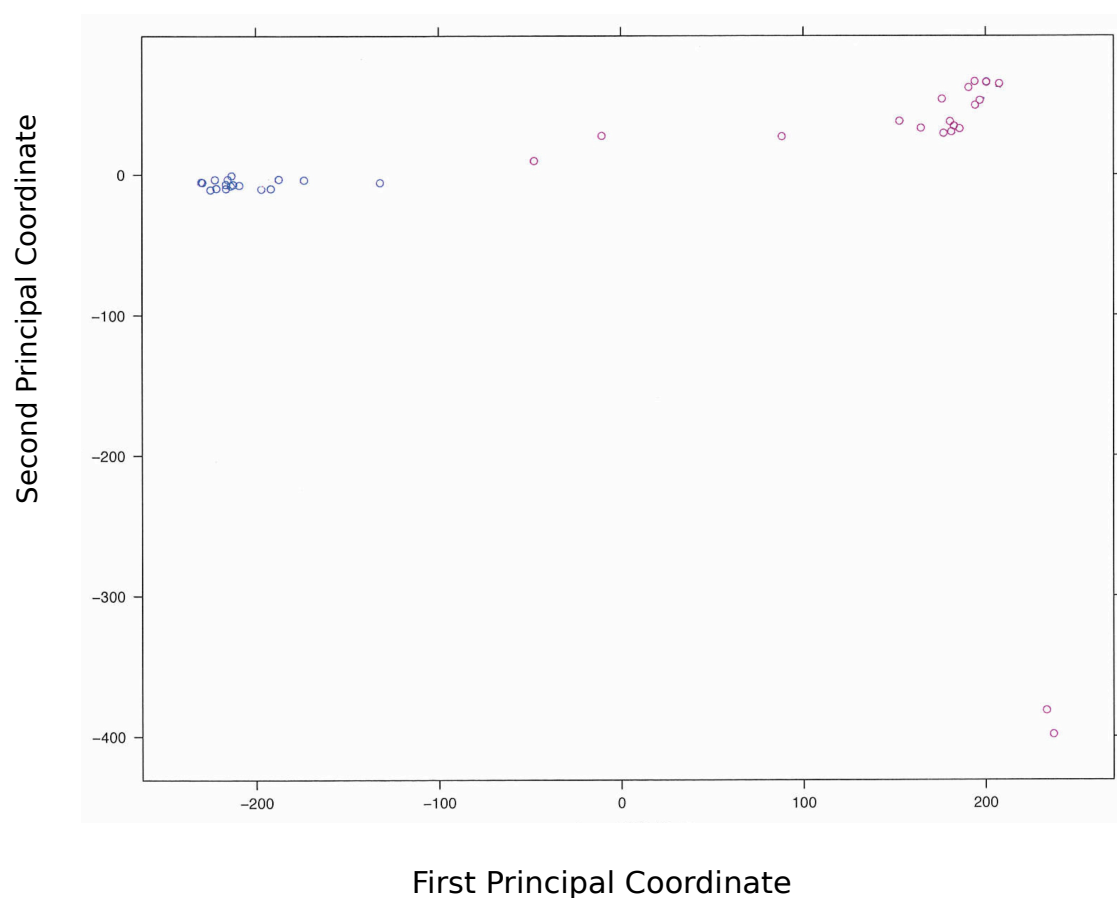
**Figure 4 . Density map of tajima D between CAM and CAP**  
More the density is high and the color tends to red

Fay and Wu's  $H$  (Fay and Wu, 2000) allows to detect mutations with high frequency and to detect positive selection, provided you have an estimator of the ancestral haplotype. Indeed, an outgroup is used to find ancestral alleles. Zeng's  $E$  (Zeng et al., 2006) is used to detect recent selective sweeps that begin to restore their polymorphism. Plots of estimates of Tajima's  $D$  against Fay & Wu's  $H$  and Zeng's  $E$  for each population are represented in Figure 5. For each population, there is a purifying selection when by  $H < 0$  (Figure 5, A & B). Finally, the  $D$  graphics can detect a recent selective sweep, stronger for the domestic population



**Figure 5 : Tajima's  $D$  against Fay & Wu's  $H$  (A & B) and Zeng's  $E$  (C & D) for CAM and CAP population**

Principal component analysis (PCA) was performed to infer affiliation of strains to each population. In a principal component analysis, the first two principal components accounted for 42 % and 34 % of the variance, respectively (Figure 2). PCA was based on 1 794 853 SNP detected within the 36 genomes. The principal component revealed two distinct clusters of samples: a cluster gathering wild samples (CAM), and a cluster gathering domestic samples (CAP).



**Figure 2 : Principal coordinate analysis among 36 samples of *Venturia inaequalis*.**

The first and second principal coordinates components for 42 % et 34 % of the variation, respectively. A group of individuals is observed according to their affiliation to a population. CAM population (pink) / CAP population (blue)

## Demographic history

We used two different methods to infer the most likely demographic model.

First, we used a model-selection approach to identify a relevant demographic model in an ABC framework. We performed coalescent simulations under various models to compute the expected distributions of summary statistics related to polymorphism and divergence patterns (Figure 1). Then, we compared the simulated statistics with the observed ones calculated from the data. We considered four models : Strict Isolation (SI), Constant Migration (CM), Ancient Migration (AM) and Secondary Contact (SC). For 3 models involving migration (CM-AM-SC), GWH (Genome Wide Heterogeneity) was tested by the comparison of two alternative introgression versions assuming either a single introgression rate shared by all loci (Si-Homo, AM-Homo, SC-Homo), or independently sampled introgression rates across loci (SI-Hetero, AM-Hetero , SC-Hetero).

Table I lists the various posterior probabilities of each model (7 scenarios). Using model choice procedures under an ABC framework, we clearly rejected scenario with stric isolation ( $P < 0.0003$ ). Models allowing ancient and constant migration had very low posterior probabilities (AM-Homo:  $P = 0.0005$  and AM-Hétéro  $P = 0.0417$ ; CM-Homo :  $P = 0.1299$  and CM-Hétéro  $P = 0.07$ ). In contrast, secondary contact model with homogeneous gene flow had high posterior probabilities ( $P = 0.8258$ ).

**Table 1. Classification of models**

Values represent the posterior probabilities of each model, with or without introgression. For CM, AM and SC models (models with introgression) it gives posterior probabilities for models with interlocus variation in effective migration rates (Hetero) or not (Homo).

	Models without introgression	Models with introgression	
		Homo	Hetero
IS	0.0003		
CM		0.1299	0.007
AM		0.0005	0.0417
SC		<b>0.8258</b>	0.0011

We examined parameter estimates under the best-fitting SC model of CAM/CAP divergence. In this analysis, the mutation rate was assumed to be constant over time and across loci, and equal to  $5 \times 10^{-8}$  bp/year. Estimates of the effective population sizes suggest an approximately 15.86 fold larger population in CAP population (102.000, 95% confidence interval [CI]: 12700 – 314000) than in CAM population (6430, 95% CI : 1630 - 80100). Our estimations indicate that the split in the ancestral population probably occurred to 207 000 years. After this divergence, the species would have evolved separately over a long period until a recent SC occurred 3220 years ago (95% CI: 673 – 227000).

**Table 2: Demographic parameters estimated using ABC under the SC-homo (Secondary Contact model with homogeneous migration)**

\*Effective population size (expressed as  $10^2$  individuals) for wild population (CAM), domestic population (CAP), and their ancestor (A).

$T_{split}^1$  : time of the split between CAM and CAP

$T_{SC}^2$  : time of the secondary contact between CAM and CAP

The 95 % highest posterior density intervals are shown into brackets.

	<i>Median</i>	<i>Mode</i>	<i>Intervals</i>
$N_{CAM}^*$	157	64,3	(16,3-801)
$N_{CAP}^*$	1040	1020	(127-3140)
$N_A^*$	2210	2890	(646-6140)
$T_{split}^1$	368000	207000	(48000-1270000)
$T_{SC}^2$	7320	3220	(673-227000)
$M_{m1}$	1,6225	0,6325	(0,0292-9,2250)
$M_{m2}$	1,4625	0,5850	(0,0640-8,7250)

We used a diffusion-based approach implemented in the software package —a—i to estimate the likelihood of our dataset in several models going from strict divergence to more complex models with introgression heterogeneity due to semipermeable barriers to reproduction. As shown in the heat-maps, there are not an excess of high or low-frequency derived alleles in our data compared with what is predicted by the model. The joint frequency spectra of

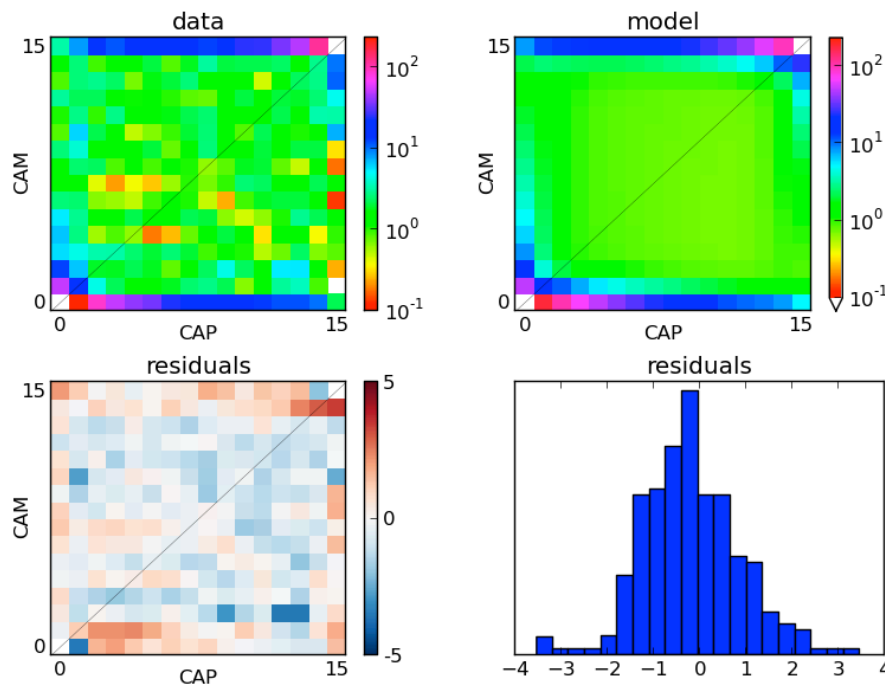
derived synonymous sites in CAM and CAP population (Fig. 5), in reference to the outgroup *V. pirina* suggests that the models of secondary contact gave the best fit to be observed two dimensional site frequency spectrum (Log Likelihood = -450.46). The other models gave only log-likelihood at least ten point under the highest log-likelihood. Likelihood ratio test with the SC model gives a  $\chi^2$  of 22 with a df of 2 that has a p-value of 0.

**Table 3:** Log-likelihood of every model tested using  $\delta a \delta i$ . The best model is indicated in bold.

Model	Log-Likelihood
SI	-470,11
IM	-461,41
IM2M	-462,48
IM2M2P	-462,37
AM	-473,24
AM2M	-463,21
AM2M2P	-463,01
PAM	-472,13
PAM2MP	-470,71
SC	-461,46
<b>SC2M</b>	<b>-450,46</b>
SC2M2P	-470,12
PSC	-461,05
PSC2M2P	-457,79

Parameters estimated from the best model, SC2M, gave the following values :  $nu_{CAM}=0.14$ ,  $nu_{CAP}=1.63$ ,  $Me_{CAM-CAP}=0.11$ ,  $Me_{CAP-CAM}=0.37$ ,  $M_{CAM-CAP}=2.72$ ,  $M_{CAP-CAM}=1.03$ ,  $T_{split}=1.71$ ,  $T_{sc}=0.27$ ,  $P=0.40$ ,  $O=0.95$ , where  $nu_x$  represents the effective size of population x relative to the ancestral population size,  $My_x$  represents the migration rate from Popx toward Popy,  $Me_{xy}$  represents the migration rate in genomic islands from Popx toward Popy,  $T_{split}$  is the splitting time,  $T_{sc}$  is the time for secondary contact, P is the proportion of genome evolving neutrally and O is the percentage of misoriented SNPs. As these estimates are given in function of a reference ancestral size. Thus assuming a divergence with *V. pirina* of 1Mya,

a mutation rate of  $5.10^{-8}$  and a one sexual generation per year, we can infer an ancestral *V. inaequalis* of *ca.* 2500 individuals. As the parameters were scaled with this ancestral population size, we can then estimate the CAM population size to be 350, the CAP one to be 4075. These two populations should have splitted 4275 generations (years) ago and went into secondary contact 675 generations or years ago. During this secondary contact, gene flow occurred at rate  $2.10^{-3}$  from CAP to CAM and  $8.10^{-4}$  in the opposite direction for neutral loci. In genomic islands, this rate is estimated to  $4.10^{-5}$  from CAM to CAP and  $1.5.10^{-4}$  in the opposite direction.



**Figure 7. Distributions of derived synonymous SNP frequencies in CAM and CAP calculated using *V. pirina* as an outgroup (best supported model : SC2M)**

To determine whether our model of secondary contact with symmetrical migration is appropriate for our data, we used metric to compare the joint allele frequency spectrum for the wild population (CAM) and domestic population (CAP) with that expected under our model: visual examination of the allele frequency spectra and calculation of the residuals between model and data. Heatmaps showing the joint allele frequency spectrum for the CAM and CAP populations compared with that expected under a secondary contact with migration model. The color of a cell at position [X,Y] in the matrix corresponds to the number of derived alleles (relative to *V. pirina*; Materials and Methods) that are at frequency X in the CAP population and frequency Y in the CAM population. The residuals represent the difference in the number of alleles predicted by the model compared with that found in the data for each bin in the spectrum (red, model predicts too many; blue, model predicts too few).

### Introgression types different according to the methods

With the two methods used to retrace divergence of *V. inaequalis*, the most likely scenario was a divergence followed by a secondary contact. However depending on the method used, either the gene flow was homogeneous or either heterogeneous. Using ABC approach, estimations indicated that homogeneous introgression ( $P=0.8258$ ) was more likely than an heterogeneous one ( $P=0.0011$ ). On the contrary, approximation diffusion model implemented in  $\delta a \delta I$  indicated a model of secondary contact with heterogeneous introgression ( $\delta a \delta I$  Log-likelihood = -450).

### **Genes involved in the reproductive isolation**

We observed more than 602 gene alignments with a differentiation index ( $F_{st}$ ) higher than 0.7. We searched orthologous genes using the Blast2GO software. Of the 602 genes, 275 orthologs were identified (2.5% of the total number of *V. inaequalis* genes).



## DISCUSSION

Genomic comparison of populations of pathogens infecting wild and domestic apple should make possible to identify the genomic zones involved in the adaptation of the pathogen to its host, and thus, the evolutionary parameters of its emergence into agroecosystems.

To identify genomic regions impacted by apples trees domestication, we sequenced the genomes of 36 strains of apple scab, *Venturia inaequalis* from domestic and wild area. We showed here that the wild and agricultural populations of *Venturia inaequalis* have experienced a recent divergence probably related to domestication of apple in Central Asia followed by a recent secondary contact. This complex situation may explain the relatively high number of genetic barriers to reproduction.

### *The most likely demographic scenario : The secondary contact*

From Approximate Bayesian Computation methods (ABC : Beaumont, 2002) and approach based on the site-frequency spectrum ( $\delta a \delta i$  : Gutenkunst, 2009), evolutionary scenario of the divergence between wild and domestic population is a scenario of secondary contact (ABC: Pvalue = 0.8258 and  $\delta a \delta i$  Log-likelihood = -450). Initially, the fungus would have followed the host plant worldwide during the domestication. More recently, gene-flow was re-established with ancestral population in Kazakhstan. According to estimates from ABC, the ancestral population splitted into two populations about 207,000 years ago (95% CI: 48,000-1,270,000). After this period of divergence, the wild and domestic populations have again been in contact there is 3220 (95% CI: 673-227000) years (Table I and II). Parameters estimates from  $\delta a \delta i$  are better in agreement with what was expected under a scenario of divergence mediated by domestication. In fact, this divergence may have happened 4000 years ago and the secondary contact six centuries ago. At now we did not have any confidence interval for these estimates. This work is on the go. Migration rates estimated by  $\delta a \delta i$  are several orders of magnitude lower than those estimated using the ABC method. This suggests a stronger isolation between both populations, probably reflecting the importance of barriers of reproductive isolation.

According to an earlier study, Gladieux (2010) showed that the posterior probabilities of divergence times between the isolated mountains of Kazakhstan's population and two other people (under an isolation with migration model: CM scenario type) suggested a divergence between these two gene pools there from 2000 to 4000 years, which coincided quite well with the period of domestication of the domestic apple (2000-7000 years, Juniper & Mabberley). In our study, the divergence time  $T_{split}$  seems too big compared in relation to genetic and historical estimates. This departure does nevertheless not challenge our scenario which has been detected by both ABC and approximate diffusion methods. The hypothesis of a potential reintroduction of apple at the foot of mountains of Kazakhstan can not however be totally validated with those times.

Our results suggest that the domestic population (CAP) is more polymorphic than the wild population (Figure 4). We can hypothesize that during the evolutionary history of domestication of apple trees from Kazakh mountains to Europe (Cornille *et al.*, 2012) and their way back to Asia, CAP strains have presumably acquired new alleles thanks to recombination. Indeed, the modern domestic apple tree arose from multiple introgressions of different species or populations of *Malus* from different regions (Cornille *et al.*, 2012). This complex evolutionary history was probably a great opportunity also for *V. inaequalis* to admix with different populations and then acquire alleles different from the original and geographically isolated populations of Kazakh mountains.

New gene flow between CAM and CAP, characterizing this secondary contact, occurred in recent period. Indeed, the period recontacting represents only 1-2% of the evolutionary history as shown by the ratio  $T_{split}/T_{sc}$  in ABC analyses. Secondary contact scenarios are not obvious to determine. The recent migration is more difficult to detect especially when that contact time is short (Roux *et al.*, 2014). It is rare indeed to distinguish between SC and IM scenario and many from errors can be committed. With a period of very recent secondary contact, the validation of the scenario SC by two independent methods, allows us to approve its robustness.

#### What type of introgression can be validated?

Depending on the method used, we distinguish two types of scenarios that differ from the type of introgression. The "ABC" method identified a scenario of secondary contact with homogeneity of gene flow while the method "δaδi" distinguishes a scenario with heterogeneous gene flow. This difference can be explained by the difference of amount of

used information between the two methods. In fact ABC used between 30 and 40 summary statistics while the diffusion-based method used 15x15 statistics of the 2D-SFS, minus the two statistics corresponding to monomorphic sites for ancestral and derived alleles respectively, that is 223 statistics. The use of such site frequency spectra is a great improvement for demographic inference (Excoffier *et al.*, 2013).

In perspective, a new analysis of ABC techniques could allow us to confirm and to refine our evolutionary scenarios. (Pudlo *et al.*, 2014). This method predicts a model and not a posterior probability. Indeed, both theoretical and algorithmic output indicate that posterior probabilities are poorly estimated by ABC. The selection of the most likely model occurs via a random forest procedure.

Introgression appears asymmetrical in the region of hybridization with a preponderance to gene flow in the direction of the domestic population CAP to the wild population CAM (Table I). If introgression seems asymmetrical, the genetic background of the wild population receives a larger amount of the domestic population allele, than the reciprocal. This hypothesis is congruent with results obtained from the PCA. Principal component analysis helped us to distinguish the relationships between the two populations. Some strains of the CAM populations are not grouped with the others, but is shifted slightly to the CAP population (Figure 2). The largest flow of population from the domestic toward the wild population is characterized by these intermediate points. Domestic population is probably be invading the wild population.

#### Potential consequences of secondary contact for pathogen populations

The effects of secondary contacts are largely ignored in pathogens. Yet through to our results and previous studies of *Venturia inaequalis* (Leroy *et al.*, 2014, Lemaire, de Gracia *et al.*, submitted), our model suggests that these situations are not occasional. As previously mentioned by Hewitt (1988), hybridization zones are real laboratories for the studies of speciations. Indeed these peculiar situations reveal genetic barriers and/or adaptive introgressions of mutations that occurred in each population (Bierne 2010; Bierne *et al.*, 2011). Conversely, hybridization between divergent populations can be an engine of diversity due to recombination generates new gene combinations in these areas, leading to hybrid swarms (Cokayne & Allan, 1926). The fertility rate and strong mutation in pathogens, allow

exploration of much faster adaptive landscape in areas of hybridization and fast access to the best fitness values than the parental populations.

Genetic recombination between divergent populations allow the transfer of genes involved in virulence, aggressiveness, dispersal potential, and consequently change the long-term epidemiological characteristics of the disease. For example, Leroy et al. (2014) suggest that the practice change cultures and globalization has created an enabling environment for see a secondary contact between pathogen populations have diverged in allopatry. For instance, European orchards were infected by migrations of virulent strains from the wild compartment (Lemaire, de Gracia *et al.*, submitted). Monteil et al. (2013) described a similar situation of migration of virulent strains of the bacteria *Pseudomonas synrigae* pv. *tomato* from non-agricultural reservoirs into crops. Thus, secondary contacts leading to hybridizations zones seems to be a real alternative to the classical origine of virulence from *de novo* mutations. This also imply that mutation toward virulences pre-existed to the introduction of resistant varieties into crops.

#### The nature of barriers and genes involved in the reproductive isolation

The existence of a secondary contact may imply the presence of barriers other than those related to adaptation to the host (Bierne *et al.*, 2011). A further study will test the hypothesis of the presence of incompatibilities also known as Dobzhansky-Muller incompatibilities (DMI, Dobzhansky, 1937; Muller, 1942; Orr, 1995) in this contact zone. This situation of secondary contact allows both to detect genes involved in adaptation to habitat but also in endogenous reproductive isolation.

Our global analysis based on all alignments of genes, allowed to detect a set of 602 genes that have a strong genetic differentiation ( $F_{st} > 0.7$ ). These 602 genes represent a set of candidate genes to study regions affected by the domestication of the apple. The identification of genes responsible for the strong differentiation, including the use of approaches based on the heterogeneity of  $F_{st}$  according to regions of the genome, allow to understand the nature of reproductive isolation barriers. However, exogenous barriers, related to the environment, are not the only ones to impede the flow of genes and to provide strong fixation indices. The so-called endogenous barriers, accumulating during the time of divergence, are also expected.

The analysis of these candidate genes will identify loci involved in genetic barriers between partially reproductively isolated populations. Identifying the evolutionary forces that characterize these barriers will be possible. Moreover, both populations CAM and CAP have also been phenotypic analysis which shows the apple domestication has profoundly changed the life history traits of *Venturia inaequalis* (de Gracia *et al.*, 2014 submitted paper in PloS One). We can also identify candidate loci potentially involved in changes in life history traits.

Human activities have produced strong changes in ecosystems on a global scale. The emergence examples of diseases on plants are particularly numerous in particular because of the profound change in the structure of the landscape that followed the domestication of plants (Stuckenberg and McDonald, 2008). The identification of factors favouring the emergence of new fungal diseases becomes absolutely necessary to prevent the adverse effects of human activities on the environment (Anderson et al. 2004).

## REFERENCES

- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004.** Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**: 535-544.
- Angiuoli SV & Salzberg SL. 2010.** Mugsy: fast multiple alignment of closely related whole genomes? *Bioinformatics* **27**: 334-342.
- Bertorelle G, Benazzo A, Mona S. 2010.** ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* **19**: 2609-2625.
- Beaumont MA, Zhang W, Balding DJ. 2002.** Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025-2035.
- Beaumont MA. 2012.** Approximate Bayesian computation in evolution and ecology. *Annual Reviews of Ecology, Evolution and Systematics* **41**: 379–406.
- Bierne N. 2010.** The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* **64**: 3254-72.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011.** The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044-2072.
- Blum MGB & Francois O. 2010.** Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*. **20**: 63–73.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005.** Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676.
- Cornille A, Gladieux P, Smulders MJ, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang XG, Tenailon MI, Giraud T. 2012.** New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* **8**: e1002703.
- Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010.** Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**: 410-418.
- Csilléry K, François O, Blum MGB. 2012.** abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology & Evolution* **3**: 475-479.

**De Gracia M, Cascales M, Expert P, Bellanger MN, Le Cam B, Lemaire C.** How did host domestication modify life history traits of its pathogens? *submitted to Plos One*

**De Mita S & M Siol. 2012.** EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics* **13**: DOI : 10.1186/1471-2156-13-27

**Dobzhansky T. 1937.** Genetics and the origin of species. Columbia University Press, New York, 364 pp.

**Dutech C, Barres B, Bridier J, Robin C, Milgroom MG, Ravigne V. 2012.** The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Molecular Ecology* **21**: 3931–3946.

**Duvaux L, Belkhir K, Boulesteix M, Boursot P. 2011.** Isolation and gene flow: inferring the speciation history of European house mice. *Molecular Ecology*. **20**: 5248-5264.

**Excoffier L, Estoup A, Cornuet JM. 2005.** Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727-1738.

**Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013.** Robust demographic inference from genomic and SNP data. *PLOS Genetics* **9**: e1003905.

**Excoffier L, Hofer T, Foll M. 2009.** Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285-298.

**Fay JC & Wu CI 2000.** Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.

**Fourcade Y, Chaput-Bardy A, Secondi J, Fleurant C, Lemaire C. 2013.** *From the cover :* Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology* **22**: 2065–2073.

**Gattepaille LM, Jakobsson M, Blum MG. 2013.** Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity* **110**: 409-419.

**Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, Didelot F, Le Cam B. 2011.** Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology* **20**: 4521-4532.

**Gladieux P, Zhang X-G, Afoufa-Bastien D, Valdebenito Sanhueza RM, Sbaghi M, Le Cam B. 2008.** On the origin and spread of the Scab disease of apple: out of central Asia. *PLoS One* **3**: e1455.

- Gladieux P, Zhang XG, Roldan-Ruiz I, Caffier V, Leroy T, Devaux M, Van Glabeke S, Coart E, Le Cam B. 2010.** Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology* **19**: 658-674.
- Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM. 2005.** Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* **102**: 1092–1097.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009.** Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5**: e1000695.
- Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. 2005.** Bayesian estimation of recent migration rates after a spatial expansion. *Genetics* **170**: 409-417.
- Harris SA, Robinson JP, Juniper BE. 2002.** Genetic clues to the origin of the apple. *Trends in Genetics* **18**: 426-430.
- Hewitt GM. 1988.** Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology & Evolution* **3**: 158-167.
- Hudson RR. 2002.** Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337-338 doi:10.1093/bioinformatics/18.2.337
- Hey J & Pinho C. 2012.** Population genetics and objectivity in species diagnosis. *Evolution* **66**: 1413-1429.
- Hey J & Nielsen R. 2004.** Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747–760.
- Hey J & Nielsen R. 2007.** Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences USA* **104**: 2785–2790.
- Jombart T, Collins C, Solymos P, Ahmed I, Calboli F, Cori A. 2013.** Adegnet: an R package for the exploratory analysis of genetic. R package version 1.4-2.
- Juniper BE, Mabberley DJ. 2006.** *The story of the apple*. Timber Press, Portland.



**Lemaire C, De Gracia M, Leroy T, Michalecka M, Lindhard Pedersen H, Gladieux P, Guerin F, a Le Cam B.** Take a walk on the wild side : on the origin of the virulent *Vf* strains of the agent of apple scab, *Venturia inaequalis*. *submitted to Plos Pathogen*

**Leroy T, Le Cam B, Lemaire C. 2014.** When virulence originates from non-agricultural hosts: new insights into plant breeding. *Infection, Genetics and Evolution*. doi:10.1016/j.meegid.2013.12.022.

**Leroy T, Lemaire C, Dunemann F, Le Cam B. 2013.** The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology* **13**:doi:10.1186/1471-2148-13-64.

**Lê Van A, Gladieux P, Lemaire C, Cornille A, Giraud T, Durel CE, Caffier V, Le Cam B. 2012.** Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications* **5**: 694-704.

**Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012.** SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**:18.10.1186/2047-217X-1-18

**Marjoram P & Tavaré S. 2006.** Modern computational approaches for analysing molecular genetic variation data. *Nat. Rev. Genet.* **7**: 759–770.

**Monteil CL, Cai R, Liu H, Mechan Llontop ME, Leman S, Studholme DJ, Morris CE, Vinatzer BA. 2013.** Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytologist* **199**: 800-811.

**Moller EM, Bahnweg G, Sandermann H, Geiger HH. 1992.** A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Research* **20**: 6115–6116.

**Morgan J & Richards A. 1993.** *The book of apples*. The Ebury Press, London.

**Muller HJ. 1942.** Isolating mechanisms, evolution and temperature. *Biological Symposia* **6**: 71–125.

**Nielsen R & Wakeley J. 2001.** Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.

**Orr HA. 1995.** The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.

- Pudlo P, Marin JM, Estoup A, Cornuet JM, Gautier M, Robert CP. 2014** ABC models choice via random forests. 28 pages, 9 figures, 4 tables. 2014. <hal-01067925>
- Richards CL, Carstens BC, Lacey Knowles L. 2007.** Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography* **34**: 1833-1845.
- Rokas A. 2009.** The effect of domestication on the fungal proteome. *Trends Genet* **25**: 60–63.
- Ross-Ibarra J, Tenaillon M, Gaut BS. 2009.** Historical divergence and gene flow in the Genus *Zea*. *Genetics* **181**: 1399-1413.
- Ross-Ibarra J, Wright SI; Foxe JP Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008.** Patterns of Polymorphism and Demographic History in Natural Populations of *Arabidopsis lyrata*. *PLoS One* 3:e2411 doi: 10.1371/journal.pone.0002411
- Roux C. 2010.** Effets de la selection naturelle et de l'histoire démographique sur les patrons de polymorphisme nucléaire. 135pp. Université de Lille I, Lille.
- Roux C, Castric V, Pauwels M, Wright SI, Saumitou-Laprade P, Vekemans X. 2011** Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS One* **6**: e26872
- Roux C, Fraisse C, Castric V., et al. 2014.** Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology* **27**: 1662-1675.
- Roux C, Pauwels M, Ruggiero MV, Charlesworth D, Castric V, Vekemans X. 2012.** Recent and ancient signature of balancing selection around the S-locus in *Arabidopsis halleri* and *A. lyrata*. *Mol Biol Evol.* **30**: 435–447.
- Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M., McDonald BA, Wang J, Schierup MH. 2011.** The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella Graminicola* and its wild sister species. *Genome Research* **21**: 2157–2166.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, KalyanaramanA, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. 2010.** The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**: 833-939.
- Wood F. 2003.** The Silk Road. The Folio Society, London.

**Zeng K, Fu YX, Wu C. 2006.** Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. *Statistic Genetics* **174** : 1431-1439.

**Zerbino DR, Birney E. 2008.** Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**: 821-829.

# Discussion générale

---

Cette thèse avait pour objectif de mieux comprendre les conséquences des activités humaines sur l'émergence des maladies. La réflexion a été alimentée par deux exemples d'émergence du champignon *Venturia inaequalis*, responsable de la tavelure du pommier ; l'un lié aux travaux de sélection variétale qui ont facilité l'introduction dans les vergers d'une population virulente préalablement présente en dehors des agrosystèmes, l'autre lié aux effets de la domestication du pommier sur les populations du champignon. Nous avons tenté de décrire l'histoire évolutive de ces populations puis de mesurer les modifications phénotypiques et de détecter les signatures génomiques liées à la transition de *V. inaequalis* d'un habitat sauvage vers les agroécosystèmes.

## 1. L'émergence de la population virulente *Vf*

L'événement initiateur de ce projet a été l'étude de l'émergence de la virulence *Vf* en Europe (Chapitre 1). La connaissance du modèle d'émergence des virulences était fondamentale avant de formuler des hypothèses quant aux mécanismes pouvant être responsables de la diversification des populations de pathogènes (écologiques ou non écologiques). Nous avons ainsi tenté de décrire l'histoire évolutive de la population porteuse de la virulence *Vf* en répondant à plusieurs questions : (1) quel est le scénario d'émergence le plus probable ? (2) de quand date(nt) ce(s) événement(s) d'émergence ? Ce travail a été réalisé en utilisant plusieurs méthodes d'assignation et un jeu de données microsatellites. Nous avons testé des scénarios en utilisant (1) plusieurs accessions collectées sur le pommier *M. floribunda*, géniteur porteur du gène *Vf* utilisé dans les croisements pour apporter le caractère résistant, (2) des populations Européennes de souches virulentes et non virulentes échantillonnées dans des vergers de *M. domestica* et enfin (3) des souches collectées sur l'ancêtre du pommier domestique *M. sieversii* dans les montagnes Kazakhes (population CAM). Nos résultats suggèrent que la virulence *Vf* était présente, à l'état sauvage, dans une population qui aurait divergé de la population présente sur hôtes sensibles depuis plusieurs milliers d'années (environ 24000 ans). Les vergers Européens auraient donc été infectés par des migrations de souches virulentes du compartiment sauvage vers les agrosystèmes. Ce résultat majeur pose question sur les modalités d'utilisation des sources de résistance introgressées dans les programmes de sélection. Les programmes d'amélioration végétale pour la résistance aux pathogènes utilisent en effet régulièrement des gènes de résistance provenant du compartiment sauvage pour l'amélioration de la résistance des espèces cultivées. Notre

étude montre que le caractère durable d'une résistance peut être éphémère si la virulence préexiste sur les espèces hôtes portant le gène introgressé. Echantillonner de manière exhaustive les populations hôtes potentiellement porteuses de virulence n'est bien sûr pas envisageable, en revanche être conscient que ce risque potentiel pourrait modifier certaines pratiques. Par exemple, le pommier *M. floribunda* est utilisé comme arbuste d'ornementation et comme pollinisateur dans les vergers. Etant à la fois source de la résistance génétique et source de la virulence, cette cohabitation a évidemment fortement favorisé l'installation de la population virulente dans les vergers.

L'utilisation du gène de résistance *Vf* dans les agrosystèmes aurait donc agi comme une passerelle pour l'invasion des habitats cultivés par la population virulente, jusqu'alors absente. Ce cas de figure est, à notre connaissance, peu décrit. Notre modèle propose donc une alternative au scénario classiquement proposé pour décrire l'émergence d'une virulence qui se fait à partir des hôtes sensibles à la suite d'une mutation favorable au locus d'avirulence du pathogène. Cette thèse montre que la virulence pourrait également émerger à partir du compartiment sauvage comme décrit par Leroy *et al.*, (2014).

Pour conclure, cette étude démontre l'importance de prendre en compte les populations de pathogènes présents sur les hôtes sauvages pour élucider l'origine de nouvelles émergences de virulence et prévoir l'évolution des maladies à long terme.

## 2. Impact de la domestication

Cette première étude, décrivant un contact secondaire possible au sein de notre pathosystème combinée aux données historiques, nous a incité à étudier l'impact des modifications de l'agriculture au sein de la zone de domestication du pommier. L'impact de la domestication sur les organismes pathogènes a rarement été étudié malgré l'importance capitale de comprendre ce processus pour envisager les conséquences à long terme. Répondre à cette problématique nous paraissait tout à fait envisageable puisque nous possédions au sein de l'équipe une collection de souches de *V. inaequalis* échantillonnées aussi bien en milieux cultivés (population CAP) qu'en milieux sauvages (population CAM). En d'autres termes nous disposions d'une population qui a subi la domestication du pommier et une autre nous servant de témoin « pré-domestication. » (Gladieux *et al.*, 2010)

A partir de ce matériel précieux, une comparaison phénotypique associée à une approche de génomique comparative nous a permis d'évaluer les changements liés à la

domestication à partir de souches isolées sur pommiers domestiques (CAP) et sur pommiers sauvages (CAM).

### 2.1. Evolution des traits d'histoire de vie: mis en évidence d'un processus de « pestification »

Dans un premier temps, nous avons procédé à plusieurs comparaisons phénotypiques pour étudier les changements des différents traits d'histoire de vie impliqués dans la dissémination du champignon de la tavelure, associées à la domestication de son hôte. Nos résultats suggèrent que le passage de l'habitat sauvage aux vergers aurait eu un impact sur les paramètres de succès d'infection sur hôte compatible et de dispersion efficace. En effet, la transition entre les deux milieux est associée à une augmentation de la taille des spores et de la capacité de sporulation. De plus, les distributions de ces deux caractères de la population domestique chevauchent avec ceux de la population sauvage. En d'autres termes, les agroécosystèmes permettraient la sélection de souches ayant une dispersion plus efficace à courte distance. Au risque de paraître finaliste, nos résultats suggèrent que la capacité d'une souche de verger à produire des spores de grande taille lui augmenterait sa probabilité à se déposer sur un hôte voisin génétiquement identique. Sachant que les souches sont déjà adaptées aux hôtes présents dans le verger (milieu caractérisé par une grande homogénéité et une densité de génotypes), ces spores ont donc une forte probabilité d'être également adaptées à ces hôtes. Dans un habitat homogène, les souches sporulant le plus et possédant les spores les plus grosses auraient donc un avantage sélectif.

La domestication des plantes est généralement associée à une forte sélection disruptive en faveur des caractéristiques de chaque environnement (Harlan *et al.*, 1973). Les résultats de notre étude contrastent avec les résultats attendus théoriques où la domestication crée une nouvelle niche d'adaptation provoquant une forte sélection des caractères qui sont ensuite associés au « syndrome de domestication ». Nous n'avons pas observé de telles adaptations *de novo* parmi les souches prélevées dans l'environnement anthropisé. En effet, la diversité phénotypique de cette population, semble bien préexister dans la population sauvage. Deux raisons pourraient expliquer l'absence de diversification phénotypique après la domestication. Contrairement à d'autres champignons tels que *Penicillium* (Ropars *et al.*, 2014) ou *Saccharomyces* (Hittinger, 2013), *V. inaequalis* n'a pas été sélectionné par l'homme pour ces caractères favorables. De plus, le passage d'un environnement à l'autre lors du suivi d'hôte est peut être trop récent pour permettre d'observer des modifications de caractères aussi importantes.

## 2.2. Barrières semi-perméables aux flux de gènes

Afin d'enrichir notre étude sur l'impact de la domestication chez *V. inaequalis*, une approche en génomique comparative a été effectuée sur la base d'un reséquencage de 36 souches provenant du centre d'origine du pathogène. A partir des méthodes ABC et celle basée sur les équations de diffusions  $\delta a d i$ , nous avons pu détecter le scénario le plus probable : un contact secondaire avec hétérogénéité de flux de gènes. Cette hétérogénéité s'expliquerait par la présence de barrières semi-perméables au flux de gènes (Roux *et al.*, 2013, 2014). L'existence d'un contact secondaire peut impliquer la présence de barrières autres que celles liées à l'adaptation à l'hôte.

## 3. Les contacts secondaires : quelles perspectives ?

Nos différentes études sur *V. inaequalis* suggèrent de nombreuses fois que les remises en contact secondaire sont des situations très probables et qu'elles ne doivent pas être mises à l'écart face aux hypothèses écologiques souvent invoquées.

L'existence de flux de gènes entre ces deux populations nous a permis de détecter par une approche en génomique comparative, les gènes impliqués dans l'adaptation de l'habitat, mais aussi dans l'isolement reproductif. Cependant, la situation de remise en contact secondaire nous amènera par la suite à interpréter ces locus selon un cadre de lecture différent : les descriptions de barrières chez le pathogène après une remise en contact secondaire laisse de grandes chances d'identifier des incompatibilités génétiques (barrières endogènes relatives à la dépression hybride) de type de Dobzhansky-Müller (DMI: Dobzhansky, 1937; Muller, 1942; Orr, 1995) accumulées au cours de la divergence et responsables, au moins en partie de l'absence de flux de gènes. Nos résultats indiquent que le fond génique de la population domestique commence à envahir la population sauvage. Après avoir comparé différents traits d'histoire de vie, quelles seraient les conséquences à l'avenir si cette introgression persistait ? La création de « supers souches » ayant un potentiel de dispersion accrue et une agressivité augmentée ? La recombinaison génétique de population divergente permet une introgression d'allèles entre les fonds génétiques. Cette situation pourrait être favorable au transfert de gènes impliqués dans la virulence, l'agressivité, le potentiel de dispersion, et par conséquent à la modification des caractéristiques épidémiologiques à long terme de la maladie.



Les exemples d'émergence de maladies sur plantes sont particulièrement nombreux en raison notamment de la profonde modification de la structure des paysages qui a fait suite à la domestication des plantes (Stuckenbrock & McDonald, 2008). La menace que font peser les champignons phytopathogènes sur les écosystèmes cultivés comme sauvages ne cesse de s'accroître depuis 20 ans (Fisher *et al.*, 2012). L'identification des facteurs favorisant l'émergence de ces nouvelles maladies fongiques devient donc absolument nécessaire afin de prévenir les conséquences néfastes des activités humaines sur l'environnement (Anderson *et al.* 2004). La durabilité des résistances aux pathogènes réside en grande partie sur une focalisation plus importante des premières étapes dans les programmes de sélection. Un effort plus grand sur la connaissance de l'aire de répartition de l'hôte géniteur de la résistance, accompagné d'une redéfinition des agrosystèmes permettraient dans un premier temps de limiter de nouvelles émergences.

## 4. Perspectives

### 4.1. A court terme

- Dans notre étude, la croissance mycélienne a été analysée en mesurant le diamètre du mycélium. Il sera envisageable de réaliser une analyse plus poussée de la structure du mycélium. Une modélisation de cette structure fractale serait plus appropriée pour évaluer la croissance du mycélium. En effet, le mycélium est formé d'une succession d'hyphes ayant des motifs de ramifications semblables (Papagianni, 2006). De plus, nos résultats suggèrent que la croissance est indépendante de la capacité à produire des spores asexuées. Or, on s'attendait plutôt à ce que la taille du mycélium soit significativement corrélée au nombre de spores produits. Cette observation suggère que la mesure de l'agressivité actuelle (pourcentage de la surface de la feuille portant des lésions) ne reflète pas la réelle dynamique de croissance de *V. inaequalis*. Sachant que le mycélium de *V. inaequalis* pousse sous la cuticule de la feuille, l'utilisation de méthodes basées sur la fluorescence de chlorophylle serait mieux appropriées pour mesurer l'agressivité (Belin *et al.*, 2014). D'après Pringle et Taylor (2002), les souches présentant des spores de grande taille ont généralement un taux de germination plus élevé. Dans notre étude, nous n'avons observé aucune corrélation entre la taille des spores et le taux de germination. Une étude de corrélation entre la taille des spores et la survie des spores germées serait cependant intéressante à étudier. En effet, une étude préliminaire récente au sein du laboratoire a montré qu'une forte mortalité des

spores peut intervenir plusieurs jours après la germination (comm pers).

- Dans ce contexte de mise en contact secondaire dans la zone d'origine du pathogène, une étude complémentaire permettra de tester l'hypothèse de l'existence d'incompatibilités de type DMI dans cette zone. Une étude en cours dans l'équipe nous laisse fortement penser que cette hypothèse est fort plausible. De nouveaux échantillonnages réalisés sous la forme de transect (habitat sauvage-habitat cultivé) ont été réalisés en août 2012 dans cette zone. Neuf sites et plus de 300 souches ont été prélevés entre Almaty et la frontière chinoise. Au cours de ma thèse, j'ai participé au traitement de ce jeu de données et identifié 181 SNP dans des gènes différenciant fortement les populations sauvages et anthropisées (fort  $F_{st}$ ). L'analyse des clines génomiques (Gompert & Buerkle, 2011) est actuellement en cours, les premiers résultats indiquent qu'il existe plusieurs cas de couplage de barrières (comm pers Lemaire). A court terme, il est prévu d'appliquer cette méthode dans les vergers où une mise en contact secondaire entre populations virulentes et populations endémiques a été mise en évidence.
- Lorsque j'ai débuté ce travail de thèse, un des premiers objectifs était de rechercher des traces de sélection sur l'ensemble des sites polymorphes du génome en s'appuyant sur les 36 génomes séquencés. Après de nombreuses tentatives d'alignements, nous avons constaté que ce travail était gigantesque et que le résultat ne serait probablement pas à la hauteur de nos attentes. Malgré une couverture de séquençage relativement élevée (50X) faisant appel à la technologie Illumina, le nombre de scaffolds moyen était trop important pour envisager de rechercher les traces de sélection par la réalisation d'une HapMap sur la base d'un alignement de 36 génomes (nombre de scaffolds moyen chez CAM : 8495 / nombre de scaffolds moyen chez CAP : 6821 ). Nous nous heurtons à la présence de zones répétées riches en Adénosine – Thymine représentant un tiers des génomes de *V. inaequalis*. Cette structure particulière en isochore a fortement complexifié l'assemblage et le travail d'alignement multiple. Des études ont montré que les DMI pouvaient impliquer des régions non codantes. L'alignement global *via* une HapMap permettrait de ne pas se limiter aux régions codantes étudiées dans cette étude. Une tentative d'assemblage est programmée en janvier 2015 à partir de lectures de séquences obtenues par la technologie PacBio dont

la taille moyenne attendue est de 8 Kb. Une meilleure qualité d'assemblage ainsi qu'une meilleure robustesse des méthodes d'alignement de génomes devraient alors nous permettre d'avoir le matériel nécessaire à la réalisation d'une Hapmap, au moins sur une grande partie des génomes et ainsi localiser les signatures de sélection en relation avec la domestication du pathogène.

- Au cours de cette étude nous avons détecté 602 gènes présentant un très fort  $F_{st}$  entre les deux populations, autant de gènes candidats pouvant faire obstacle aux flux de gènes homogènes entre les deux populations. Il serait également intéressant de pouvoir localiser des gènes candidats potentiellement impliqués dans les changements de traits d'histoire de vie. La maîtrise dans l'équipe de la transformation du pathogène par recombinaison homologue permettra d'effectuer des séries de transformations et donc d'interchanger des allèles entre les deux populations. Des échanges d'allèles aux locus responsables de barrières permettraient de décortiquer les bases génétiques de l'isolement reproducteur de ce pathogène.

#### 4.2. *A plus long terme*

« *V. inaequalis* : vers un modèle d'étude / de référence ? »

A ce jour, nous présentons l'une des premières études de l'impact de la domestication d'un hôte sur le pathogène associé, combinant une approche phénotypique et génomique. Introduit par Saleh et al. (2014), le processus de pestification, qui correspond à l'évolution des caractères liés à la fitness chez un champignon pathogène provoquée par la domestication de son hôte, a été décrit. Comme l'évolution de ces traits d'histoire de vie est fortement dépendante de chaque agent pathogène, nous ne pouvons pas d'ores et déjà étendre nos conclusions à d'autres patho-systèmes. Toutefois, nos résultats mettent l'accent sur la menace potentielle des populations d'agents pathogènes vivant dans un environnement sauvage pour la durabilité de la résistance dans les cultures (Lê Van *et al.*, 2012; Monteil *et al.*, 2013; Leroy *et al.*, 2014). L'identification et la caractérisation des gènes candidats détectés au travers de l'histoire complexe de *V. inaequalis* liée à la domestication, pourraient permettre à l'équipe de faire de la tavelure un modèle de référence dans cette thématique dans les années à venir.

# Bibliographie

---

- Agrios GN. 2005.** Plant Pathology. San Diego, CA: Elsevier.
- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004.** Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**: 535-544.
- Aylor DE. 1998.** The aerobiology of apple scab. *Plant Disease* **82**: 838-848.
- Balter M. 2007.** Seeking agriculture's ancient roots. *Science* **316**: 1830-1835.
- Barbash DA, Siino DF, Tarone AM, Roote J. 2003.** A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proceedings of the National Academy of Sciences USA* **100**: 5302-5307.
- Bates MR, Buck KW, Brasier CM. 1993.** Molecular relationships of the mitochondrial DNA of *Ophiostoma ulmi* and the NAN and EAN races of *O.novo-ulmi* determined by restriction fragment length polymorphisms. *Mycological Research* **97**: 1093-1100.
- Beaumont MA & Balding DJ. 2004.** Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**: 969-980.
- Beaumont MA & Nichols RA. 1996.** Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**: 1619-1626.
- Beaumont MA, Zhang W, Balding DJ. 2002.** Approximate Bayesian Computation in population genetics. *Genetics* **162**: 2025-2035.
- Belin E, Rousseau D, Boureau T, Caffier V. 2013.** Thermography versus chlorophyll fluorescence imaging for detection and quantification of apple scab. *Computers and Electronics in Agriculture* **90**: 159-163.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011.** The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044-2072.
- Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D. 2007.** Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biology* **5**: e236.
- Brasier CM. 1995.** Episodic selection as a force in fungal microevolution with special reference to clonal speciation and hybrid introgression. *Canadian Journal of Botany* **73**: 1213-1221.

- Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. 2006.** Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*. **314**: 1292-1295.
- Buckler ES, Thornsberry JM, Kresovich S. 2001.** Molecular diversity, structure and domestication of grasses. *Genetical Research* **77**: 213-218.
- Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, *et al.* 2007.** Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* **3**: 1745–1756.
- Campbell CS, Greene CW, Dickinson TA. 1991.** Reproductive-biology in subfam *Maloideae* (*Rosaceae*). *Systematic Botany* **16**: 333-349.
- Carisse O, Meloche C, Boivin G, Jobin T. 2009.** Action Thresholds for Summer Fungicide Sprays and Sequential Classification of Apple Scab Incidence. *Plant Diseases*. **93**: 490-8.
- Chou JY, Hung YS, Lin KH, Lee HY, Leu JY. 2010.** Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biology* **8**: e1000432.
- Cornille A, Gladieux P, Smulders MJ, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L *et al.* 2012.** New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* **8**: e1002703.
- Cornille A, Giraud T, Bellard C, Tellier A, Le Cam B, Smulders MJ, Kleinchmit J, Roldan-Ruiz I, Gladieux P. 2013.** Postglacial recolonization history of the European crab apple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* **22**: 2249-2263.
- Cornille A, Giraud T, Smulders MJ, Roldan-Ruiz I, Gladieux P. 2014.** The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**: 57-65.
- Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Notteghem JL & Kohn LM. 2005.** Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* **170**: 613-630.
- Crosby JA, Janick J, Pecknold PC, Korban SS, O’Connon PA, Ries SM, Goffreda J, Voordeckers A. 1992.** Breeding apples for scab resistance: 1945–1990. *Fruit Varieties Journal* , **46**: 145–166.

- Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010.** Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**: 410-418.
- Darwin C. 1859.** The origin of species. London, UK: John Murray.
- Darwin C. 1882.** The variation of animals and plants under domestication. London, UK: John Murray.
- Diamond J. 2002.** Evolution, consequences and future of plant and animal domestication. *Nature* **418**: 700-707.
- Dobzhansky T. 1937.** Genetics and the origin of species. Columbia University Press, New York, 364 pp.
- Dunemann F, Gläss R, Bartsch S, Saad Eldin M, Peil A, Bus V. 2012.** Molecular cloning and analysis of apple *HcrVf* resistance gene paralogs in a collection of related species. *Tree Genetics & Genomes* **8**: 1095-1109.
- Ferree PM & Barbash DA. 2009.** Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biology* **7**: e1000234.
- Fisher MC, Henk D, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012.** Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484** : 186–94.
- Flor HH. 1971.** Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **9**: 275-278.
- Foll M & Gaggiotti O. 2008** A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**: 977-993.
- Forsline PL. 1995.** Adding diversity to the national apple germplasm collection: collecting wild apples in Kazakhstan. *New York Fruit Quarterly*, 3, 3-6.
- Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC. 2002.** Collection, maintenance, characterization and utilization of wild apples of Central Asia. In: *Horticultural Reviews* pp. 1-61. John Wiley & Sons, Inc.
- Forsline PL, Dickson EE, Luby JL. 1994.** Collection of wild *Malus*, *Vitis* and other fruit species genetic resources in Kazakhstan and neighbouring republics. *Horticultural Reviews* **29**: 2-61.

- Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. 2006.** Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics* **38**: 953-956.
- Gessler C & Pertot I. 2012.** *Vf* scab resistance of *Malus*. *Trees-structure and function* **26**: 95-108.
- Giraud T, Gladieux P, Gavrillets S. 2010.** Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution* **25**: 387-395.
- Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, Didelot F, Le Cam B. 2011.** Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology* **20**: 4521-4532.
- Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B. 2008.** On the origin and spread of the scab disease of apple: out of Central Asia. *PloS One* **1**: e1455.
- Gladieux P, Zhang XG, Roldan-Ruiz I, Caffier V, Leroy T, Devaux M, Van Glabeke S, Coart E, Le Cam B. 2010.** Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology* **19**: 658-674.
- Glémin S & Bataillon T. 2009.** A comparative view of the evolution of grasses under domestication. *New phytologist* **183**: 273-290.
- Gomez-Alpizar L, Carbone I, Ristaino JB. 2007.** An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proceedings of the National Academy of Sciences USA* **104**: 3306-3311.
- Gompert Z, Buerkle CA. 2011.** Bayesian estimation of genomic clines. *Molecular Ecology* **10**: 2111-2127.
- Gu Z, David L, Petrov D, Jones T, Davis RW, Steinmetz LM. 2005.** Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* **102**: 1092-1097.
- Guerin F and Le Cam B. 2004.** Breakdown of the scab resistance gene *Vf* in apple leads to a founder effect in populations of the fungal pathogen *Venturia inaequalis*. *Phytopathology* **94**: 364-369.
- Guerin F, Gladieux P, Le Cam B 2007.** Origin and colonization history of newly virulent



strains of the phytopathogenic fungus *Venturia inaequalis*. *Fungal Genetics and Biology* **44**: 284-292.

**Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009.** Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5**: e1000695.

**Harlan J, de Wet J, Price E. 1973.** Comparative evolution of cereals. *Evolution* **27**: 311-325.

**Harris SA, Robinson JP, Juniper BE. 2002.** Genetic clues to the origin of the apple. *Trends in Genetics* **18**: 426-430.

**Hewitt GM. 1988.** Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology & Evolution* **3**: 158-167.

**Hittinger CT. 2013.** *Saccharomyces* diversity and evolution: a budding model genus. *Trends in Genetics* **29**: 309-317.

**Holb IJ. 2006.** Effect of six sanitation treatments on leaf litter density, ascospore production of *Venturia inaequalis* and scab incidence in integrated and organic apple orchards. *European Journal of Plant Pathology* **115**: 293-307

**Holb I, Heijne B, Withagen J, Jeger M. 2004.** Dispersal of *Venturia inaequalis* ascospores and disease gradients from a defined inoculum source. *Journal of Phytopathology* **152**: 639-646.

**Hossain MB, Piotrowski M, Lensing J, Gau AE. 2009.** Inhibition of conidial growth of *Venturia inaequalis* by the extracellular protein fraction from the antagonistic bacterium *Pseudomonas fluorescens* Bk3. *Biological Control* **48**: 133-139.

**Hough LF. 1944.** A survey of the scab resistance of the foliage on seedlings in selected apple progenies. *Proceedings of the American Society for Horticultural Science* **4**: 260-272.

**Ioos R, Andrieux A, Marçais B, Frey P. 2006.** Genetics characterization of the natural hybrid *Phytophthora alni* as inferred from nuclear and mitochondrial DNA analyses. *Fungal Genetics and Biology* **43**: 511-529.

**Janzen DH. 1980.** When is it coevolution. *Evolution* **34**: 611-612.

- Jha G, Thakur K, Thakur P. 2009.** The *Venturia* apple pathosystem: pathogenicity mechanisms and plant defense responses. *Journal of Biomedicine and Biotechnology* **2009**, 680160.
- Juniper BE & Mabberley DJ. 2006.** The story of the apple. *Timber Press*, Portland.
- Kilian B, Ozkan H, Walther A, Kohl J, Dagan T, Salamini F, Martin W. 2007.** Molecular diversity at 18 *loci* in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (Einkorn) domestication: implications for the origin of agriculture. *Molecular Biology and Evolution* **24**: 2657–2668.
- Korban S & Skirvin R. 1984.** Nomenclature of the cultivated apple. *Horticultural Science* **19**: 177-180.
- Kruuk LE, Baird SJ, Gale KS, Barton NH. 1999.** A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**: 1959–1971.
- Le Cam B, Parisi L, Arene L. 2002.** Evidence of two formae speciales in *Venturia inaequalis*, responsible for apple and *Pyracantha* scab. *Phytopathology* **92**: 314-320.
- Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. 2008.** Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**: 1065-1073.
- Leroy T, Le Cam B, Lemaire C. 2014.** When virulence originates from non-agricultural hosts: new insights into plant breeding. *Infection, Genetics and Evolution*. doi:10.1016/j.meegid.2013.12.022.
- Leroy T, Lemaire C, Dunemann F, Le Cam B. 2013.** The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology*. **13**:doi:10.1186/1471-2148-13-64.
- Leuenberger C & Wegmann D. 2009.** Bayesian computation and model selection without likelihoods. *Genetics* **184**: 243-252.
- Lê Van A, Gladieux P, Lemaire C, Cornille A, Giraud T, Durel CE, Caffier V, Le Cam B. 2012.** Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications* **5**: 694-704.

- Lewontin RC & Krakauer J. 1973.** Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* **74**: 175-195.
- Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006.** Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences USA* **103**: 9578–9583.
- McDonald BA & Linde C. 2002.** Pathogen population genetics, evolutionary potential and durable resistance. *Annual Review of Phytopathology* **40**: 349-379.
- McHardy WE. 1996.** *Apple scab: biology, epidemiology and management*. The American Phytopathological Society USA: St Paul.
- McHardy WE, Gadoury DM, Gessler C. 2001.** Parasitic and biological fitness of *Venturia inaequalis*: relationship to disease management strategies. *Plant Disease* **85**: 1036-1051.
- Maheshwari S & Barbash DA. 2011.** The genetics of hybrid incompatibilities. *Annual Review of Genetic*. **45**: 331-355.
- Malek J. 1993.** The cat in ancient Egypt. *London: British Museum Press*. 144p.
- Maynard Smith J & Haigh J. 1974.** The hitch-hiking effect of a favourable gene. *Genetical Research* **23**: 23-35.
- Menon R. 1956.** Studies on *Venturiaceae* on Rosaceous plants. *Phytopathologische Zeitschrift* **27**: 117-146.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009.** A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**: 373-375.
- Mills WD & Laplante AA. 1951.** Diseases and insects in the orchard. *Cornell Extension Bulletin*, 711.
- Monteil CL, Cai R, Liu H, Mechan Llontop ME, Leman S, Studholme DJ, Morris CE, Vinatzer BA. 2013.** Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytologist* **199**: 800-811.
- Moran NA & Jarvik T. 2010.** Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624-7.
- Morgan J & Richards A. 1993.** The book of apples. *The Ebury Press, London*.

- Morris CE, Bardin M, Kinkel LL, Moury B, Nicot PC, Sands DC. 2009.** Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. *PloS Pathogens* **5**: e1000693.
- Muller HJ. 1942.** Isolating mechanisms, evolution and temperature. *Biological Symposia* **6**: 71–125.
- Munkacsi AB, Stoxen S, May G. 2008.** *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. *Proceedings of the Royal Society* **275**: 1037-1046.
- Nielsen R & Wakeley J. 2001.** Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.
- Orr HA. 1995.** The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.
- Papagianni M. 2006.** Quantification of the fractal nature of mycelial aggregation in *Aspergillus niger* submerged cultures. *Microbial Cell Factories* **5**: doi:10.1186/1475-2859-5-5.
- Parisi L, Durel CE, Laurens F. 2000.** First report on the presence of *Venturia inaequalis* race 7 in french apple orchards. In: *Integrated control of pome fruit disease* (ed. Parisi L), pp. 99-104, Gent
- Parisi L, Laurens F, Guillaumes J, Kruger J. 1993.** A new race of *Venturia inaequalis* virulent to apples with resistance due to the *Vf* gene. *Phytopathology* **83**: 533-537
- Payseur BA & Nachman MW. 2000.** Microsatellite variation and recombination rate in the human genome. *Genetics* **156**: 1285-1298.
- Phadnis N & Orr HA. 2009.** A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* **323**: 376-379.
- Phipps JB, Robertson K.R., Rohrer J.R., Smith P.G. 1991.** Origins and evolution of subfam *Maloideae* (*Rosaceae*). *Systematic Botany* **16**: 303-332.
- Pickersgill B. 2007.** Domestication of plants in the Americas: insights from mendelian and molecular genetics. *Annals of Botany* **100**: 925-940.

- Potter D, Eriksson T, Evans RC. 2007.** Phylogeny and classification of *Rosaceae*. *Plant Systematics and Evolution* **266**: 5-43.
- Presgraves DC, Balagopalan L, Abmayr SM, Orr HA. 2003.** Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423**: 715-719.
- Pringle A, Taylor JW. 2002.** The fitness of filamentous fungi. *Trends in Microbiology* **10**: 474-481.
- Raabe, RD & Gardner, MW. 1972.** Scab of *Pyracantha*, *Loquat*, *Toyon*, and *Kageneckia*. *Phytopathology* **62**: 914-916.
- Robinson JP, Harris SA, Juniper BE. 2001.** Taxonomy of the genus *Malus* Mill.(*Rosaceae*) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Systematics and Evolution* **226**: 35-58.
- Ropars J, López-Villavicencio M, Dupont J, Snirc A, Gillot G, Coton M, Jany JL, Coton E, Giraud T. 2014.** Induction of sexual reproduction and genetic diversity in the cheese fungus *Penicillium roqueforti*. *Evolutionary Applications* **7**: 433-441.
- Rokas A. 2009.** The effect of domestication on the fungal proteome. *Trends in Genetics* **25**: 60–63.
- Rosenthal BM. 2008.** How has agriculture influenced the geography and genetics of animal parasites? *Trends In Parasitology* **25**: 67-70.
- Roux C. 2010.** Effets de la selection naturelle et de l’histoire démographique sur les patrons de polymorphisme nucléaire. 135pp. Université de Lille I, Lille.
- Roux C, Fraisse C, Castrie V., et al. 2014.** Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology* **27**: 1662-1675.
- Roux C, Tsagkogeorga G, Bierne N, Galtier N, 2013.** Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution* **30**: 1574–1587.
- Salamini F, Ozkan H, Brandolini A, Schâfer-Pregl R, Martin W. 2002.** Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics* **3**: 429-441.

- Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D. 2014.** South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, *Magnaporthe oryzae*. *New Phytologist* **201**: 1440-1456.
- Salvaudon L, Héraudet V, Shykoff JA. 2008.** *Arabidopsis thaliana* and the Robin Hood parasite: a chivalrous oomycete that steals fitness from fecund hosts and benefits the poorest one? *Biology Letters* **4**: 526-529.
- Sawamura K & Yamamoto MT. 1997.** Characterization of a reproductive isolation gene, zygotic hybrid rescue, of *Drosophila melanogaster* by using minichromosomes. *Heredity* **79**: 97–103.
- Serpell JA. 2000.** Domestication and history of the cat, *In: The domestic cat: the biology of its behaviour*. Turner D.C. & Bateson P. (Eds), Cambridge: Cambridge University Press, 180-192.
- Sivanesan A. 1977.** *The taxonomy and pathology of Venturia species*. J. Cramer, Vaduz.
- Smereka KJ, Machardy WE, Kausch AP. 1987.** Cellular differentiation in *Venturia inaequalis* ascospores during germination and penetration of apple leaves. *Canadian Journal of Botany* **65**: 2549-2561.
- Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. 2007.** Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution* **24**: 398-411.
- Stukenbrock EH, MacDonald BA. 2008.** The origins of plant pathogens in agroecosystems. *Annual Review of Phytopathology* **46**: 75-100.
- Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M., McDonald BA, Wang J, Schierup MH. 2011.** The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella Graminicola* and its wild sister species. *Genome Research* **21**: 2157–2166.
- Stukenbrock EH, Christiansen FB, Hansen TH, Dutheil JY, Schierup MH. 2012a.** Fusion of two divergent fungal individuals led to the recent emergence of a new widespread pathogen species. *Proceedings of the National Academy of Sciences USA*. doi: 10.1073/pnas.1201403109
- Stukenbrock EH, Quaedvlieg W, Javan-Nikah M, Zala M, Crous P, MacDonald P. 2012b.** *Zymoseptoria ardabilia* and *Z. pseudotritici*, two progenitor species of the *septoria*

*tritici* leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). *Mycologia* doi: 10.3852/11-374.

**Tang S & Presgraves D.C. 2009.** Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* **323**: 779-782.

**Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997.** Inferring coalescence times from DNA sequence data. *Genetics* **145**: 505-518.

**Thrall PH & Burdon JJ. 1999.** The spatial scale of pathogen dispersal: consequences for disease dynamics and persistence. *Evolutionary Ecology Research* **1**: 681-701.

**Torriani SF, Brunner PC, McDonald BA. 2011.** Evolutionary history of the mitochondrial genome in *Mycosphaerella* populations infecting bread wheat, durum wheat and wild grasses. *Molecular Phylogenetics and Evolution* **58**: 192-197

**Van Valen L. 1973.** A new evolutionary law. *Evolutionary Theory* **1**: 1-30.

**Vavilov NI. 1930.** Wild progenitors of the fruit trees of Turkistan and the Caucasus and the problem of the origin of fruit trees. *International Horticultural Congress* Group B: 271-286.

**Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. 2010.** The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**: 833-939.

**Vitalis R, Dawson K, Boursot P. 2001.** Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**: 1811–1823.

**Wood F. 2003.** *The Silk Road*. The Folio Society, London.

**Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005.** The effects of artificial selection on the maize genome. *Science* **308**: 1310–1314.

**Zaffarano PL, MacDonald BA, Linde CC. 2008.** Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution* **62**: 1418-36.

# Thèse de Doctorat

Marie DE GRACIA

## Génomique évolutive de l'agent pathogène de la tavelure du pommier, *Venturia inaequalis*, dans le cadre de la domestication de son hôte

*Evolutionary genomics of the pathogen of apple scab, Venturia inaequalis, as part of the domestication of its host*

### Résumé

Les changements phénotypiques et génétiques consécutifs au processus de domestication et plus largement du passage d'un environnement sauvage aux agro-écosystèmes sont bien décrits chez les espèces végétales. Cependant, l'impact de ces changements de traits d'histoire de vie sur les populations de pathogènes infectant les hôtes domestiqués a jusqu'alors été très peu abordé. En particulier, quelles sont les conséquences sur les traits d'histoire de vie des pathogènes du passage d'un milieu sauvage caractérisé par une grande hétérogénéité d'hôtes à des agro-écosystèmes bien plus denses et homogènes génétiquement? Est-ce que le génome des pathogènes porte des signatures de changements démographiques ou de sélection en relation avec la domestication de leurs hôtes? Ces questions traitées dans le cadre de la présente thèse ont porté sur le champignon *Venturia inaequalis*, agent pathogène de la tavelure du pommier, qui partage en Asie Centrale le même centre d'origine que son hôte endémique sauvage *Malus sieversii*. Les travaux reposent principalement sur 1) l'étude du passage de souches sauvages virulentes sur pommier résistant *Vf* expliquant le contournement rapide de cette résistance et 2) l'étude de génomique des populations et de variations phénotypiques comparant une population de *V. inaequalis* échantillonnée sur *M. sieversii* au Kazakhstan, préalablement identifiée comme étant une relique de la population ancestrale, et d'une population Kazakhe de milieu anthropisé. Ces deux populations sont génétiquement différenciées et partiellement isolées reproductivement. L'utilisation des deux méthodes d'inférence l'une l'ABC et l'autre dadi basées sur l'alignement de 10 994 gènes échantillonnés au sein des deux populations révèlent une histoire complexe de contact secondaire, où le champignon aurait dans un premier temps « suivi » la plante hôte à travers le monde au cours de la domestication (modèle de l'host tracking) puis échangerait à nouveau depuis peu des gènes avec la population ancestrale au Kazakhstan. Cette situation originale de remise en contact secondaire favorisant la mise en évidence d'incompatibilités génétiques, permet à la fois d'envisager de détecter des gènes impliqués dans l'adaptation à l'habitat (pommier cultivé versus *Malus sieversii*) et dans l'isolement reproductif post-zygotique. L'analyse des 36 génomes séquencés a ainsi permis d'identifier plus de 602 gènes présentant un indice de différenciation (*Fst*) supérieur à 0,7 autant de gènes verrous faisant potentiellement obstacle aux flux de gènes homogènes entre ces deux populations. Enfin, l'analyse phénotypique de ces deux populations montre que la domestication du pommier a profondément modifié les traits d'histoire de vie de *V. inaequalis* liés à sa dispersion. L'ensemble de ces analyses a donc permis d'identifier des locus candidats potentiellement impliqués dans des modifications des traits d'histoire de vie, et dans des barrières géniques, en lien avec la domestication du pommier.

### Mots clés

Zone de domestication, Emergence, Traits d'histoire de vie, Introgression, *Vf*, barrières génétiques, *Venturia inaequalis*, Contact secondaire.

### Abstract

Phenotypic and genetic changes occurring during the process of domestication are well described in plants. However, the impact of domestication in life history traits of their pathogens has been poorly studied. In particular, what are the consequences on the life history traits of pathogens that switch from the wild habitat characterized by a high host genetic and spatial heterogeneity to a much more dense and genetically homogeneous agro-ecosystems? Have pathogen's genomes particular signatures of demographic changes or of selection related to the domestication of their host? Here we focused on the fungus *Venturia inaequalis*, causal agent of apple scab, that shares in Central Asia the same center of origin of its wild endemic host *Malus sieversii*. In the first part of this work we retrace evolutionary history of a population from the wild that was responsible for the rapid breakdown of the *Vf* resistance gene in apple orchards. We then highlight the threat of wild habitat to scab resistance apple cultivars and thus the necessity to take into account the wild in the management of resistance genes. In the second part, a comparative study based on phenotypic and genomic data was carried out between two populations sampled in Kazakhstan on *M. sieversii*, either in anthropized areas or in non anthropized area, the latter being identified as a relic of the ancestral population of *V. inaequalis*. These two populations were genetically differentiated and partially reproductively isolated. The use of two methods of inference (ABC and dadi) based on the alignment of 10 994 genes sampled in the two populations revealed a complex history of a secondary contact event. The fungus would have first followed its host worldwide by "Host-tracking" during the domestication process and then very recently it would re-exchange genes with its ancestral population in Kazakhstan. Secondary contact context is particularly favorable to detect genetic incompatibilities, this particular situation could then allow identification of genes involved in adaptation to habitat (cultivated apple versus *M. sieversii*) and in post-zygotic reproductive isolation. Analysis of 36 sequenced genomes has identified more than 602 genes with an index of differentiation (*Fst*) greater than 0.7, what represents numerous candidates genes of potential barriers to homogeneous gene flow between these two populations. Comparative phenotypic analysis between these two populations such spores size and capacity to sporulate showed that apple domestication would have also modified life history traits of *V. inaequalis* related to its dispersion. This study has allowed identification of candidate loci potentially involved in changes in life history traits, and genetic barriers in pathogen related to the domestication of apple.

### Key Words

Domestication zone, Emergence, life history traits, Introgression, *Vf*, genetic barriers, *Venturia inaequalis*, secondary contact.